



Universidade Federal da Paraíba
Centro de Ciências Humanas, Letras e Artes
Programa de Pós-Graduação em Linguística

Acurácia de um Modelo Fonotático de Entropia Máxima Aplicado ao Português Brasileiro

Fernando Cabral Alves

Orientador: Prof.º Dr. Rubens Marques de Lucena

João Pessoa

2017



Universidade Federal da Paraíba
Centro de Ciências Humanas, Letras e Artes
Programa de Pós-Graduação em Linguística

Área de Concentração: Teoria e Análise Linguística
Linha de Pesquisa: Diversidade e Mudança Linguística

Acurácia de um Modelo Fonotático de Entropia Máxima aplicado ao Português Brasileiro

Fernando Cabral Alves

Dissertação de Mestrado
apresentada ao Programa de Pós-
Graduação em Linguística –
PROLING – da Universidade
Federal da Paraíba (UFPB), como
requisito parcial para obtenção do
grau de Mestre em Linguística.

Professor Orientador: Dr. Rubens Marques de Lucena

João Pessoa

2017

A474a Alves, Fernando Cabral.

Acurácia de um modelo fonotático de entropia máxima
aplicado ao português brasileiro / Fernando Cabral Alves. -
João Pessoa, 2017.

118 f.: il. -

Orientador: Rubens Marques de Lucena.
Dissertação (Mestrado) - UFPB/ CCHL

UFPB/BC

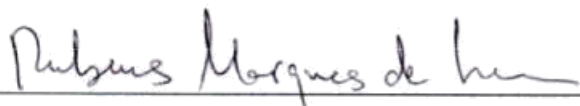
CDU: 801(043)

Acurácia de um Modelo Fonotático de Entropia Máxima aplicado ao Português Brasileiro

Fernando Cabral Alves

Dissertação aprovada em 29 de março de 2017

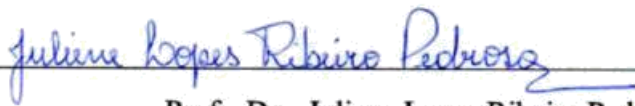
Banca Examinadora:



Prof. Dr. Rubens Marques de Lucena
(Orientador)

Profa. Dra. Ana Carla Estellita Vogeley

Membro avaliador



Profs. Dra. Juliene Lopes Ribeiro Pedrosa

Membro avaliador

João Pessoa

2017

Às grandes mulheres em
minha vida: Fernanda, Raísa,
Socorro.

AGRADECIMENTOS

À Fernanda Lúscia Cabral de Araújo, minha mãe, pelo apoio diário sem o qual eu não teria condições de escrever este trabalho e por aturar minha chatice durante a elaboração do mesmo. Ao meu pai, Severino Alves de Souza, também pelo apoio e compreensão quando precisei.

À Raísa de Sousa Claro, minha noiva, por haver dedicado horas e horas ao longo de alguns dias (mesmo estando muito ocupada) gravando e regravando áudios para o experimento descrito no Capítulo II, e aguentando reclamações de um produtor exigente. Agradeço também por dividir meu tempo com a ciência, compreendendo meus momentos de reclusão e me incentivando a fazer trabalhos como este.

A Rubens Marques de Lucena, pelos seis anos de orientação e pelo tratamento humano durante todo esse período. Agradeço por todas as oportunidades acadêmicas, aulas, conselhos, discussões, revisões de texto, puxões de orelha, elogios, humildade, confiança e por diversos outros fatores de uma rica relação intelectual e humana sem a qual este trabalho não existiria.

Aos membros da banca, Ana Vogeley e Juliene Pedrosa, pelo forte incentivo ao novo (algo raro na academia), e pelo tempo e esforço dedicados para fazer contribuições que efetivamente melhoraram o texto final.

A Bruce Hayes e Colin Wilson, os autores do modelo título da dissertação, pela troca de e-mails que permitiu esclarecer questões de ordem técnica não discutidas na publicação (HAYES&WILSON, 2008). As informações foram essenciais para uma real compreensão de alguns componentes do modelo, como será observado no apêndice A, e também na implementação de nossa aplicação ao português brasileiro.

A todos os autores que consultei na elaboração desta dissertação. Este trabalho é tanto meu quanto deles.

Ao Programa de Pós-graduação em Linguística da Universidade Federal da Paraíba, em especial aos professores que me lecionaram: Dermeval da Hora, Leonardo Wanderley Lopes, José Ferrari Neto, Evangelina Brito de Faria.

Ao servidor Valberto Cardoso, pelo brilhante apoio com a parte burocrática da pós-graduação ao longo dos dois anos em que estive no programa.

A Ribamar de Castro, meu eterno mestre, por me encorajar nos primeiros passos em fonologia e por despertar uma motivação irreversível.

A Pedro Felipe de Lima Henrique, meu amigo e colega de pesquisa, pelas lúcidas e produtivas conversas sobre fonologia, linguística e ciências em geral, e por me encorajar a realizar este trabalho quando o projeto ainda era uma vaga ideia.

Ao meu amigo Leandro da Costa Santos, pelo interesse no trabalho, o que me permitiu falar sobre o mesmo para alguém que não é da área, e também por compartilhar das dificuldades e prazeres de fazer ciência, tornando a jornada mais leve.

RESUMO

O presente trabalho faz parte dos estudos que buscam representar e investigar sistemas linguísticos a partir de modelos matemáticos. Neste contexto, um modelo de Entropia Máxima para fonotática e aprendizado fonotático desenvolvido por Hayes e Wilson (2008) apresentou uma alta correlação com dados experimentais quando aplicado ao inglês, superando outras propostas de modelagem fonotática (HAYES&WILSON, 2008, p.401). Porém, apesar dos bons resultados, desconhecemos qualquer trabalho no Brasil que utilize o modelo em questão ou modelos de Entropia Máxima em geral. Uma vez que o modelo é universal (i.e. aplicável a qualquer língua), tomamos como objetivo desta dissertação medir o nível de acurácia do modelo quando aplicado ao português brasileiro (doravante PB). O texto se divide em três capítulos. No primeiro, tratamos de descrever em detalhes o modelo a ser testado. No segundo, apresentamos a metodologia utilizada para: i) aplicar o modelo ao português brasileiro; e ii) coletar dados experimentais para medir a acurácia das previsões do modelo obtidas em i). Os procedimentos metodológicos envolveram a criação de dois softwares, um para transcrição fonológica automática do PB e outro para realização de experimentos de estimação de magnitude. Por fim, no terceiro capítulo, apresentamos os resultados. Nas duas aplicações realizadas, a correlação entre as previsões do modelo e os dados experimentais, medida a partir do coeficiente Pearson, ficaram em torno de 0 e 0,5, demonstrando assim uma dependência linear muito mais fraca que aquela encontrada para o inglês (0,946).

Palavras-chave: Fonotática. Entropia Máxima. Aprendizado de Máquina. Estimação de Magnitude.

ABSTRACT

The present work is part of the studies that seek to represent and investigate linguistic systems using mathematical models. In such context, a Maximum Entropy model of phonotactics developed by Hayes and Wilson (2008) has exhibited a high level of accuracy in relation to experimental data when applied to English, outperforming other phonotactic modelling proposals. Nevertheless, despite its good results, we are ignorant of any work in Brazil which makes use of the model or of Maximum Entropy models in general. Since the model is universal (i.e. applicable to any language), we have taken our objective to be measuring the level of accuracy of the model when applying it to Brazilian Portuguese. The text is divided into three chapters. In the first chapter, we have described in details the model to be tested. In the second one, we have presented the methodology employed to: i) apply the phonotactic model to Brazilian Portuguese; and ii) collect experimental data against which we measure the accuracy of the model predictions obtained in i). The methodological procedures involved the creation of two softwares, one for automated phonological transcription of Brazilian Portuguese and a second one for carrying out magnitude estimation experiments. Finally, in chapter three we show the results. In two applications, the correlation between model predictions and experimental data, measured by the Pearson coefficient, were found to be in the region of 0 and 0,5, thus showing a much weaker linear dependence than that found for English (0,946).

Keywords: Phonotactics. Maximum Entropy. Machine Learning. Magnitude Estimation.

LISTA DE TABELAS

Tabela 01 – Espaço total de restrições em relação a n , para $ N_1 = 97$ e $ N_2 = 90$	39
Tabela 02 – Aplicação completa – Médias dos logs das estimações: F.V. - F.C.....	85
Tabela 03 - Aplicação dos onsets – Médias dos logs das estimações: F.V. - F.C	85
Tabela 04 – Correlações da aplicação completa.....	85
Tabela 05 – Correlações da aplicação dos onsets.....	87

LISTA DE GRÁFICOS

Gráfico 01 - função $H(x) = -[x \log_2(x) + (1 - x) \log_2(1 - x)]$, com $x \in [0. 1]$...	29
Gráfico 02 – Aplicação completa: médias das razões F.V. - F.C.....	83
Gráfico 03 - Aplicação dos onsets - médias das razões: F.V. - F.C.	84
Gráfico 04 – Distribuição de $E_C_{razão}$ em relação a M_C	86
Gráfico 05 – Distribuição de $E_O_{razão}$ em relação a M_O	88

LISTA DE QUADROS

Quadro 01 – Conjunto de traços distintivos postulados para GU.....	34
Quadro 02 - Maxent PL em Linguagem Algorítmica.....	43
Quadro 03 – Consoantes do PB em traços distintivos.....	45
Quadro 04 – Vogais do PB em traços distintivos.....	46
Quadro 05 – Processos utilizados no transcritor automático.....	51
Quadro 06 – Alterações no alfabeto fonológico.....	54
Quadro 07 – Parâmetros das aplicações	62
Quadro 08 – Amostra da aplicação completa: restrição, F.V., F.C., e valor Maxent da F.V.	64
Quadro 09 – Amostra da aplicação dos <i>onsets</i> : restrição, F.V., F.C., e valor Maxent da F.V.	65

LISTA DE FIGURAS

Figura 01 – Funcionalidade do Maxent PL	19
Figura 02- Funcionalidade do transcritor automático.....	50
Figura 03 – Interface de usuário do software do Maxent PL	56
Figura 04 - Amostra de dados para aplicação dos onsets	61
Figura 05 – Programa de estimação de magnitude de retas (tela inicial da interface do usuário)	74
Figura 06 – Programa de estimação de magnitude de retas (interface do usuário para estimação).....	75
Figura 07 – Output da estimação de retas do participante 01	76
Figura 08 – Programa de estimação de magnitude fonotática (tela inicial da interface do usuário)	77
Figura 09 – Programa de estimação de magnitude fonotática (interface do usuário para estimação).....	79
Figura 10 – Output da estimação fonotática do participante 03	80

SUMÁRIO

INTRODUÇÃO	13
CAPÍTULO I: O Modelo de Entropia Máxima de Hayes e Wilson (2008).....	22
1. A Fonotática no Maxent PL	23
1.1. Os pesos.....	25
1.1.1 O Princípio de Entropia Máxima e a Regra de Delimitação	26
1.1.2 O cálculo dos pesos	30
1.2 As restrições	33
1.2.1 O espaço total de restrições	38
1.2.2 Seleção de restrições: acurácia e generalidade	40
CAPÍTULO II: Metodologia	44
2. Objetivos metodológicos	44
2.1. Construindo <i>M</i>	44
2.1.1 O inventário fonológico em traços distintivos	44
2.1.2 A amostra de dados	47
2.1.3 Parâmetros do software	55
2.1.4 Resultados das aplicações	58
2.1.5 Determinando <i>S</i> e <i>M</i>	63
2.2. Construindo <i>E</i>	67
2.2.1 Estimação de Magnitude	69
2.2.2 Estimação de Magnitude aplicada à fonotática	72
2.2.2.1 Participantes.....	72
2.2.2.2 Materiais	72
2.2.2.3 Procedimentos	73
2.2.2.3.1 Etapa de calibragem	73
2.2.2.3.2 Determinando <i>E</i>	77
CAPÍTULO III: Resultados	83
CONSIDERAÇÕES FINAIS.....	83
REFERÊNCIAS	91
ANEXO A1	94
ANEXO A2	96
ANEXO B1	101
ANEXO B2	103
APÊNDICE A	106
APÊNDICE B.....	110
APÊNDICE C1.....	111
APÊNDICE C2.....	111
APÊNDICE D	112
APÊNDICE E1	113
APÊNDICE E2	115
APÊNDICE E3.....	117

A mathematical model is never a completely accurate representation of a physical situation— it is an *idealization*. A good model simplifies reality enough to permit mathematical calculations but is accurate enough to provide valuable conclusions. It is important to realize the limitations of the model. In the end, Mother Nature has the final say.

(James Stewart – Calculus: Early Transcendentals)

INTRODUÇÃO

O presente trabalho faz parte dos estudos que buscam representar e investigar sistemas linguísticos a partir de modelos matemáticos. Neste contexto, nosso objetivo é medir o nível de acurácia de um modelo universal (i.e. aplicável a qualquer língua natural) de fonotática desenvolvido por Hayes e Wilson (2008) quando aplicado ao português brasileiro (doravante PB). Escolhemos desenvolver este trabalho em específico por dois principais motivos: i) desconhecemos outros trabalhos no Brasil que utilizem modelos de Entropia Máxima, tal como o de Hayes e Wilson (2008); ii) em uma aplicação com o inglês, este modelo em particular apresentou um alto nível de acurácia em relação a dados experimentais, superando outras propostas testadas com o mesmo conjunto de dados empíricos (HAYES&WILSON, 2008, p.401).

Para transformar nosso objetivo em um problema bem definido (e.g. definindo acurácia rigorosamente), contudo, precisamos começar a introduzir a perspectiva matemática do modelo e o fenômeno que o motiva. Esta introdução buscará, portanto, apresentar o material necessário para que, ao final dela, possamos definir precisamente o problema de pesquisa ao qual toda a dissertação é dedicada.

Nosso ponto de partida será a definição de *língua (formal)* advinda da Teoria das Línguas Formais (HOPCROFT; ULLMAN; MOTWANI, 2001[1979]; CHISWELL, 2009). Chamemos de *alfabeto* um conjunto qualquer de símbolos, denotado por A , e de *palavras*, o conjunto A^n , com $n \in \mathbb{N}$, contendo todas as sequências finitas $a_1 a_2 \dots a_n$ com $a_i \in A$, e $i \in n$. Formalmente, essas sequências podem ser definidas como membros do produto cartesiano de $A_1 \times A_2 \dots \times A_n$, escritos sem os parênteses usuais.

É importante mencionar que $A^0 := \{\varepsilon\}$, onde ε é a chamada *palavra vazia*, que possui as seguintes propriedades: $|\varepsilon| = 0$ (cardinalidade/comprimento igual a zero), $\varepsilon p = p$ (identidade sob concatenação)¹. Note-se também que $\varepsilon \neq \emptyset$. Podemos, agora, apresentar as definições de *universo* e de *língua*:

Definição: um *universo* A^* de um *alfabeto* A qualquer é dado por $A^* = \bigcup_{n \geq 0} A^n$.

Definição: uma *língua* L sobre um *alfabeto* A é um subconjunto de A^* .

¹ A operação de concatenação será definida mais abaixo.

Embora propostas mais recentes em Linguística Matemática expandam a definição acima para obter uma álgebra de signos (relacionando significados aos elementos da *língua* por meio de operações) (ver KRACHT, 2003), veremos que a definição puramente sintática que apresentamos é suficiente para trabalhar com nosso modelo de interesse e, consequentemente, para toda a dissertação.

Sendo o conceito intuitivo do que venha a ser um *fonema*, assim como certas propriedades deste objeto, consideravelmente bem estabelecidas em linguística, o tomaremos aqui como primitivo. Isso nos deixa na posição confortável de ignorar a difícil e extensa tarefa de definir *fonema* a partir de conceitos mais básicos e alguns axiomas². Como nosso objetivo aqui não é construir os fundamentos matemáticos da fonologia, tal empreendimento nos distanciaria de nosso verdadeiro objetivo.

Feitas as considerações dos dois últimos parágrafos, chamaremos de *inventário fonológico* o *alfabeto* I formado pelos *fonemas* de uma dada língua natural³, e de *fonologia segmental* qualquer *língua* S_I sobre uma união do inventário I com outro *alfabeto* de símbolos não terminais (e.g. coda, onset, sílaba)⁴. Assim como no nível sintático, a fonologia, enquanto disciplina empírica, não está interessada em qualquer S_I . Busca-se construir a *língua* que corresponda o máximo possível à língua natural que se deseja representar. Esse subconjunto do *universo*, como sabemos, deve ser potencialmente infinito (uma propriedade comumente chamada pelos linguistas de “infinitude virtual” ou “infinitude digital”), daí surge o problema ao qual grande parte da linguística formal vem se dedicando⁵. Como explica Kornai, “A primary concern of mathematical linguistics is to effectively enumerate those sets of words, sentences, etc., that play some important linguistic role. Typically, this is done by means of generating the set in question [...]” (2007, p. 9).

Para construir o subconjunto de interesse (em diferentes níveis linguísticos) a partir de um número finito de representações, a linguística formal tem tradicionalmente utilizado uma “gramática”. Para definir “gramática”, no sentido que desejamos, comecemos com um *sistema de reescrita*:

² Para uma proposta como esta, ver Kornai (2007, pp.24-26).

³ Acharmos desnecessário postular, no corpo do texto, a existência de uma bijeção dos fonemas para símbolos escritos.

⁴ Um alfabeto de símbolos não-terminais será definido mais abaixo.

⁵ Na verdade, não apenas a infinitude virtual motiva a construção das gramáticas que veremos, mas também a aceitação do postulado 2.1 que veremos mais abaixo.

Definição: um *sistema de reescrita* é um subconjunto de $A^* X A^*$.

Sejam $T = t_1 \dots t_m$ e $U = u_1 \dots u_n$ seqüências em A^* , a concatenação entre T e U , denotada TU , é definida por $t_1 \dots t_m u_1 \dots u_n$, ou seja, um elemento de A^{m+n} onde os índices de 1 até m da seqüência são $t_1 \dots t_m$ e aqueles de índice $m + 1$ até $m + n$ são $u_1 \dots u_n$. De posse dessa operação binária, deixemos o par ordenado (α, β) ser um elemento de um *sistema de reescrita*, então dizemos que para qualquer v e $w \in A^*$, $v\alpha w$ reescreve para $v\beta w$, usualmente denotado por $\alpha \longrightarrow \beta$. Agora podemos definir uma “derivação”:

Definição: se $\alpha = t_1$ e $\beta = t_m$, com α e $\beta \in A^*$, e existir uma seqüência t_1, t_2, \dots, t_m , também em A^* , tal que $t_i \longrightarrow t_{i+1}$ para todo i no intervalo $1 \leq i \leq m - 1$, escrevemos $\alpha \xrightarrow{\bullet} \beta$, e dizemos que t_1, t_2, \dots, t_m é uma *derivação* de β a partir de α .

Definimos agora uma “gramática” e uma *língua* gerada por tal “gramática”:

Definição: uma *gramática* é uma quádrupla (A_n, A_t, R, P) , em que:

- i) A_n e A_t são *alfabetos*, de maneira que $A_n \cap A_t = \emptyset$;
- ii) R é um *sistema de reescrita* finito sobre $A_n \cup A_t$, ou seja, é um subconjunto de $((A_n \cup A_t)^* X (A_n \cup A_t)^*)$;
- iii) $P \in A_t$.

Definição: uma *língua* L_G gerada por uma dada *gramática* G é definida pelo conjunto:

$$L_G = \{w \in A_t^* \mid P \xrightarrow{\bullet} w\}$$

Em Teoria das Línguas formais já foi demonstrado que classes específicas de *gramáticas*, a chamada ‘hierarquia de Chomsky’, são reconhecidas por diferentes tipos de “máquinas” (i.e. as máquinas determinam se uma dada seqüência (*string*) pertence à língua gerada pela gramática). Para uma discussão detalhada da ‘hierarquia de Chomsky’ e para as demonstrações das relações entre as gramáticas e as máquinas teóricas supracitadas, referimos os leitores para Chiswell (2009, capítulo 1). Nosso interesse aqui é apenas utilizar as definições realizadas até então para formalizar dois postulados que

são aceitos em teorias de linguística formal que utilizam *gramáticas* desde pelo menos a publicação de *Syntactic Structures* (CHOMSKY, 2002 [1957]) mas que, por algum motivo, nunca vimos sendo explicitados de maneira rigorosa:

Postulado 1: para toda *língua formal* L_G sobre o alfabeto A , se L_G é identificada com uma dada língua natural, então L_G é um conjunto infinito.

Como já vimos, este primeiro postulado motiva o trabalho com *gramáticas* e a investigação de máquinas teóricas que resolvam o problema de pertencimento ao conjunto L_G . Porém, a principal motivação para a construção de *gramáticas*, como entenderemos melhor mais abaixo, advém do seguinte postulado:

Postulado 2.1: para toda *língua formal* L_G sobre o alfabeto A , se L_G é identificada com uma dada língua natural, então $A^* \neq L_G$ (ou, equivalentemente, $A^* \subset L_G$)

Se chamarmos os elementos s em A^* que não estão em L_G de *sequências mal-formadas* e chamarmos os elementos de L_G de *sequências bem-formadas* e aceitarmos os postulados acima, podemos provar um Teorema que é geralmente assumido como verdadeiro pelos linguistas que trabalham nas teorias em questão (i.e. teorias que trabalham com *gramáticas* no sentido que definimos):

Teorema: se L_G é uma *língua formal* sobre o alfabeto A identificada com uma língua natural qualquer, então existem elementos *bem-formados* e *mal-formados* em A^* e apenas eles.

Demonstração: pelo postulado 1, sabemos que $L_G \neq \emptyset$, e, que, de fato, existem infinitos elementos *bem-formados*. Contudo, embora seja verdade que para toda sequência β , se $\beta \in L_G$, então $\beta \in A^*$, é falso afirmar que se $\beta \in A^*$, então $\beta \in L_G$, pois isso seria equivalente a dizer que $A^* = L_G$, o que contradiria o postulado 2.1. Assim, confirmamos também a existência de elementos *mal-formados* em A^* . Agora se supormos, pelo contrário, que haja um elemento $\alpha \in A^*$ que não seja *bem-formado* e não seja *mal-formado*, então $\alpha \notin A^*$, que é um absurdo. \square

Podemos dizer que este Teorema caracteriza o que chamamos de teorias categóricas em linguística. A intuição por trás do postulado 2.1 é de que embora certos objetos simbólicos existam na mente de um falante-ouvinte de uma língua natural, nem todas as permutações (estruturas) formadas a partir do sequenciamento desses símbolos fazem parte dessa língua. Partindo desse postulado, é natural assumir também que a mente desses falantes-ouvintes distingue elementos *bem-formados* daqueles *mal-formados* em relação à sua língua. Essa é a ideia que motiva os chamados testes de gramaticalidade (CHOMSKY, 1965), que nada mais são que julgamentos de falantes-ouvintes de uma dada língua acerca de uma forma linguística a fim de eliciar (já que se assume o postulado 2.1) se a forma faz parte de sua língua ou não (voltaremos a discutir tais testes na seção 2.2, quando tratarmos de nosso experimento).

Nosso interesse em explicitar e formalizar alguns dos fundamentos das teorias categóricas em linguística formal é mostrar que, embora Hayes e Wilson (2008) não tragam essa discussão, o modelo dos autores, assim como outros modelos baseados em probabilidade, difere fundamentalmente dos modelos categóricos. Assim, tentaremos aqui, mesmo que brevemente, oferecer uma proposta de formalização do que acreditamos ser postulados que se encontram na base de algumas teorias probabilísticas em linguística para que possamos compreender o que de fato o modelo de Hayes e Wilson (2008) busca modelar. Outra razão para fazer isso é que os trabalhos em linguística formal no Brasil são majoritariamente categóricos, o que pode fazer com que a ideia de gradiência na boa-formação (a qual apresentaremos mais abaixo) seja vista de maneira suspeita ou prontamente rejeitada como inútil (o que de fato é verdade quando estamos trabalhando dentro de uma teoria que aceita o postulado 2.1) se não compreendermos como os postulados e definições são diferentes para o outro grupo de teorias.

A partir de agora, para fins de abreviação, batizaremos e nos referiremos ao modelo de Entropia Máxima para fonotática e aprendizado fonotático de Hayes e Wilson (2008) como Maxent PL (nome escolhido tendo em mente o termo em inglês **Maximum Entropy Phonotactic Learner**). Propomos que o Maxent PL abandona o postulado 2.1, e o substitui (implicitamente) pelo seguinte:

Postulado 2.2: para toda *língua formal* L_G sobre o alfabeto A , se L_G é identificada com uma dada língua natural, $A^* = L_G$.

Como o leitor já deve haver concluído, uma consequência imediata do novo postulado é que só há elementos *bem-formados* em A^* , ou, conversamente, não há elementos *mal-formados* em A^* . Também não há mais problema em enumerar os elementos de L_G ⁶ (o Postulado 1 é descartado, bastando exigir que A não seja vazio), retirando assim a principal motivação para construir qualquer *gramática* do tipo que definimos. Se antes podíamos haver definido a ‘fonotática’ como uma *gramática* que determina uma língua L_G sobre a união de um *inventário fonológico* e um *alfabeto não terminal*, isso significa que não há mais necessidade para construir uma (*gramática*) *fonotática* como tal, ou até mesmo propor um *alfabeto não terminal*.

Não havendo mais necessidade de postular um *alfabeto não terminal*, podemos dizer que nossa nova *língua* corresponde às chamadas formas de superfície em linguística, assim modificamos também a definição de *fonologia segmental* que determinamos antes:

Definição: *fonologia segmental* é uma *língua* sobre um *inventário fonológico*.

A proposta, contudo, não é simplesmente definir uma língua como um “vale-tudo simbólico”, mas substituir a visão de uma fonotática como uma *gramática* que distingue binariamente (*bem-formado* x *mal-formado*) os elementos de A^* , pelo seguinte:

Definição: uma *língua* A^* sobre um *alfabeto* A é um *espaço amostral* Ω ;

Definição: a *fonotática* de uma *língua* A^* é uma *distribuição de probabilidade* sobre A^* .

Definição: a *boa-formação* de um elemento de A^* é o valor atribuído a ele pela *fonotática*.

A motivação por trás de modelos que aceitam o postulado 2.2 (em detrimento de 2.1) e também as três últimas definições acima (o Maxent PL sendo um deles) advém de resultados de diversos estudos experimentais. Como explicam Hayes e Wilson:

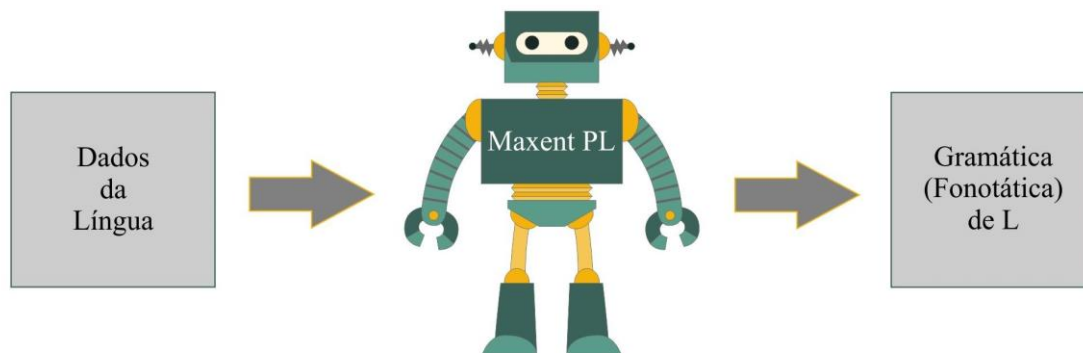
⁶ Um pouco mais tecnicamente, não há necessidade também de considerar um alfabeto com símbolos não-terminais.

In the particular domain of phonotactics, gradient intuitions are pervasive: they have been found in every experiment that allowed participants to rate forms on a scale (e.g. Greenberg e Jenkins (1964), Ohala e Ohala (1986), Coleman e Pierrehumbert (1997), Vitevitch et al. (1997), Frisch, Large, e Pisoni (2000), Treiman et al. (2000), Bailey e Hahn (2001), Hay, Pierrehumbert, e Beckman (2003), Coetzee (2004), Hammond (2004), Berent et al. (2007). (2008, p.382)

Uma discussão sobre a necessidade ou não de modificar os fundamentos de modelos padrões em linguística formal não nos interessa aqui. Entretanto, nossos resultados finais podem oferecer insights para discussões dessa natureza. Isto posto, podemos, finalmente, começar a definir nosso problema de pesquisa de maneira precisa.

Iniciamos ilustrando abaixo o funcionamento geral do Maxent PL:

Figura 01 – Funcionalidade do Maxent PL



De certa maneira, poderíamos dizer que o Maxent PL toma como entrada os dados de uma língua natural L qualquer, e, produz, como saída, uma *fonotática* para L . Contudo, na prática, o que obtemos não é exatamente a *fonotática*, pois esta, como definimos, é um conjunto infinito. O algoritmo produz um conjunto de “restrições” associadas a pesos numéricos com o qual é possível calcular a *boa-formação* de qualquer forma no *universo* L^* ⁷. Chamaremos este conjunto de *gramática fonotática*, mas só entenderemos precisamente seus componentes (i.e. restrições e pesos) no capítulo 1, quando descrevermos o Maxent PL em detalhes.

Por hora, o que nos interessa é saber apenas que ao aplicarmos os dados do PB ao Maxent PL, obteremos um objeto (i.e. uma *gramática fonotática* do PB) que determinará

⁷ Como veremos no capítulo 1, também não é diretamente a *boa-formação* que calcularemos, mas um valor proporcional a ela, chamado de *score* (pontuação).

a *boa-formação* de qualquer forma da língua⁸. A partir disto, chamemos de S um subconjunto finito de *palavras* do PB*, $\{s_1, \dots, s_n\}$, e de G uma gramática fonotática do PB obtida após uma aplicação do Maxent PL. Denotamos agora como $P_G(s_i)$ a *boa-formação* de uma forma $s_i \in S$ determinada por G e construímos o seguinte vetor:

$$M = (P_M(s_1), \dots, P_M(s_n))$$

Assumamos agora que conseguimos eliciar, a partir de um experimento, a *boa-formação* atribuída por falantes-ouvintes nativos de PB às *palavras* em S , denotando tais valores, de maneira análoga, como $P_E(s_i)$ para todo $s_i \in S$ e construindo o vetor E tal que:

$$E = (P_E(s_1), \dots, P_E(s_n))$$

Assim, chegamos ao ponto em que podemos definir precisamente o problema ao qual toda a dissertação é dedicado:

Problema de pesquisa: qual o coeficiente de correlação Pearson entre M e E ?

Basicamente, o que chamamos de “nível de acurácia” no objetivo geral descrito no primeiro parágrafo deste texto está agora definido matematicamente como o coeficiente de correlação Pearson, e as variáveis que serão comparadas são basicamente o conjunto de valores previstos pelo Maxent PL para um dado fenômeno (i.e. *boa-formação* de um conjunto de formas fonológicas do PB) e os valores experimentais do fenômeno modelado. Claramente, a solução do problema segue diretamente da construção de M e E . Portanto, a maior e principal parte de nosso trabalho se dá nos procedimentos metodológicos.

Isto posto, a presente dissertação se organiza da seguinte maneira para tratar do problema de pesquisa acima: no capítulo 1, descreveremos o Maxent PL em detalhes. No capítulo 2, descreveremos a metodologia utilizada para obtermos os valores de M e de E .

⁸ A nota anterior se aplica novamente.

Por fim, no capítulo 3, apresentaremos o resultado da correlação (e algumas outras estatísticas) juntamente com alguns comentários.

CAPÍTULO I – O Modelo de Entropia Máxima de Hayes e Wilson (2008)

Ao assumir o postulado 2.2 e as definições de *língua* como espaço amostral, *fonotática*, e *boa-formação* que vieram em seguida, o trabalho de um modelo fonotático de uma *língua* L qualquer, é, portanto, definir a *fonotática* de L . O Maxent PL representa uma dessas propostas, e, além disso, pode ser aplicado a qualquer *língua*. Neste capítulo, descreveremos em detalhes como o Maxent PL define e constrói a *fonotática* de uma língua L .

Antes de prosseguir, contudo, abrimos um parêntese neste parágrafo para realizar algumas observações importantes. Um modelo estatístico é classificado como sendo de Entropia Máxima quando obedece dois princípios matemáticos (os quais descreveremos em 1.1.1) que, por serem puramente formais, não o restringe a uma interpretação em particular. Em outras palavras, o uso de modelos de Entropia Máxima não está restrito à fonotática, à linguística, ou a qualquer ciência empírica em particular. Para outro exemplo de aplicação em linguística, sugerimos o trabalho de Goldwater e Johnson (2003), no qual um modelo de Entropia Máxima ao aparato da Teoria da Otimalidade (PRINCE&SMOLENSKY, 2004 [1993]) para aprender e modelar gramáticas com e sem variação. Para mais exemplos e uma discussão acerca da utilização desses modelos em linguística computacional, ver o trabalho de Berger, Pietra, e Pietra (1996), que introduz modelos de Entropia Máxima à área. Isto posto, pedimos ao leitor que generalize os dois princípios que apresentaremos em 1.1.1 em termos de nossos objetos específicos (i.e. restrições) para qualquer tipo de objeto estatístico que possa ser classificado e contado (e que a literatura em Entropia Máxima chama de *feature*⁹).

⁹ Também evitamos falar em *feature* para não confundir com os traços distintivos em fonologia. Para uma definição formal de *feature*, ver Klein e Manning (2003). Outra possível confusão terminológica que pode acontecer com linguistas

1 – A fonotática do Maxent PL

Começemos com um espaço amostral Ω identificado com uma *fonologia segmental*. A *boa-formação* de qualquer elemento $x \in \Omega$ é definida em uma *fonotática* P como:

$$P(x) = \frac{P^*(x)}{Z}$$

onde

$P^*(x)$ é o valor de Entropia Máxima da forma x , e
 Z é uma constante que corresponde a soma de todos os valores de Entropia Máxima de todas as formas em Ω , incluindo $P^*(x)$.

O *valor de Entropia Máxima* de uma forma x é dado por:

$$P^*(x) = \exp(-h(x))$$

onde

$h(x)$ é a pontuação¹⁰ de uma forma x

A *pontuação*, por sua vez, é:

$$h(x) = \sum_{i=1}^N w_i C_i(x)$$

onde

w_i é o peso da i ésima restrição, e
 $C_i(x)$ é o número de vezes que a forma x viola a i ésima restrição.

¹⁰ Tradução da forma em inglês ‘score’.

Unindo todas as equações e formalizando Z , podemos expressar $P(x)$ de maneira mais explícita:

$$P(x) = \frac{\text{Exp}[-(\sum_{i=1}^n w_i(x)C_i(x))]}{\sum_{y \in \Omega} \text{Exp}[-(\sum_{i=1}^n w_i(y)C_i(y))]}$$

A partir dessas equações, podemos deduzir informações importantes sobre nosso modelo. Primeiramente, como Ω é infinito, “[...] the probability of any one given form will be extremely small. What is important is the differences between these probabilities, which (as we will show) can be large and meaningful.” (HAYES&WILSON, 2008, p. 383). Além disso, sendo Z uma constante, $P(x)$ é proporcional a $P^*(x)$; logo, podemos seguir a seguinte proposta: “[...] the actual computed probabilities, while embodying the most direct interpretation of a maxent grammar, are not crucial in predicting relative well-formedness. For this reason, we will illustrate here only the calculation of scores and maxent values.” (HAYES&WILSON, 2008, p.384). Na prática, portanto, buscaremos obter apenas as *pontuações* das formas em Ω .

Como podemos ver mais claramente na última equação acima, os parâmetros fundamentais para determinar a *pontuação* e $P(x)$ são dois componentes que chamamos de restrições¹¹ e pesos (associados a essas restrições). As restrições, no Maxent PL, correspondem a restrições de marcação em linguística (McCARTHY, 2002) e os pesos nada mais são que números reais positivos associados às restrições. Não é difícil notar a partir da equação de *pontuação* que quanto maior o número de violações incorridas por uma forma, menor será a *boa-formação* da mesma, assim estabelecendo matematicamente o que esperamos intuitivamente da interação desses conceitos. É oportuno lembrar que chamamos o conjunto de restrições associadas a seus pesos de *gramática fonotática*, e, podemos entender agora por que dissemos antes que, de posse dela, é possível “[...] calcular a *boa-formação* de qualquer forma em PB*”.

Resta-nos, portanto, explicar como o Maxent PL determina o conjunto de restrições e seus respectivos pesos a partir dos dados de uma determinada língua. De fato, somente quando fizermos isso é que estaremos caracterizando realmente nosso modelo de interesse. Com o que apresentamos até então, o Maxent PL se confunde com qualquer

¹¹ Muito embora não se utilize diretamente as restrições no cálculo de *pontuação* e sim o número de vezes que uma forma viola cada restrição, isso é um resultado que segue trivialmente e depende fundamentalmente do que são e quais são as restrições.

outro modelo log linear de *fonotática*, pois ainda não apresentamos o Princípio de Entropia Máxima que faz dele um modelo Maxent (. Assim, para concluir nossa descrição do Maxent PL, dedicamos uma subseção aos pesos e outra ao conjunto de restrições, respectivamente.

1.1 Os pesos

Suponhamos, por hora, que já possuímos um conjunto de restrições. Para ilustrar, determinemos um conjunto C para o PB, contendo as seguintes restrições $\{ *pf, *pr \}$, sendo $\{ p, f, r \}$ um subconjunto do *inventário fonológico* de PB. Do ponto de vista mais geral, puramente probabilístico, restrições são chamadas de traços (do inglês *feature*¹²), e são definidas como funções indicadoras. Aqui, para nossos objetos linguísticos, uma *restrição* será definida formalmente como uma *palavra*; por exemplo, $*pf$ é igual a (p, f) e $*pr$ é igual a (p, r) . Já vimos que qualquer *palavra* em PB é dada por (a_1, \dots, a_n) com a_i pertencendo ao *inventário fonológico*. Com esta definição, podemos definir $C_i(x)$ (o número de violações de uma restrição que uma palavra x incorre e que usamos na equação de *pontuação*) para o caso específico de $*pf$ da seguinte maneira (note que usaremos uma função indicadora mas que ela não é identificada como a *restrição* em si):

$$*pf(x) = \sum_{j=1}^n I(x), \text{ com } n = |x|$$

Onde a função $I(x)$ será dada por:

$$I(x) = \begin{cases} 1, & \text{se } [(\exists a_j \in x \wedge a_j = p) \wedge (\exists a_{j+1} \in x \wedge a_{j+1} = f)] \\ 0, & \text{caso contrário} \end{cases}$$

Para formalizar $C_i(x)$ para *restrições* com 1 elemento ou mais de 2 elementos, bastaria alterações triviais nas condicionais de $I(x)$. Sabemos que a definição acima poderia ser mais eficiente computacionalmente, pois para *restrições* de 2 elementos como $*pf$, se $I(x)$ retorna o valor 1 para um a_j qualquer, não há necessidade de avaliar $I(x)$

¹² Essa não é uma boa escolha quando estamos trabalhando com objetos linguísticos, pois como sabemos, *feature* em fonologia tem outro significado bem estabelecido.

novamente para a_{j+1} , pois já sabemos que nesse caso $a_{j+1} = f$, e retornará o valor 0. Porém, queremos evitar definições mais extensas e deselegantes do ponto de vista notacional. A partir de agora assumimos que uma função como $*pf(x)$ exista para qualquer $C_i \in C$.

Como pode ser facilmente deduzido, a partir das definições de *restrição* e $C_i(x)$ que acabamos de apresentar, se o conjunto C de *restrições* já foi construído, os valores $C_i(x)$ serão então fixados, restando, como esperado, determinar apenas os pesos w_i para termos um modelo fonotático bem definido. Supondo que nosso modelo já determinou um conjunto C de *restrições* (embora só veremos como o Maxent PL faz isso na próxima subseção), tal como $\{*pf, *pr\}$ ou qualquer outro, explicaremos agora como os *pesos* são atribuídos aos elementos de C . O aprendizado dos pesos é justamente o que define o Maxent PL como um modelo de Entropia Máxima.

1.1.1 – A Regra de Delimitação e o Princípio de Entropia Máxima

Acabamos de ver, a partir da definição de $C_i(x)$, que a violação de uma *restrição* como $*pf$ se dá exatamente quando a sequência de *fonemas* pf faz parte de uma forma. Portanto, uma ideia que segue naturalmente é a de que restrições que são violadas com frequência nos dados da língua devam receber pesos menores e vice-versa; pois, como já vimos, violação de restrições com pesos maiores diminuem mais a probabilidade de uma forma do que violação de restrições com pesos menores; logo, se desejamos que a *boa-formação* signifique de fato o que gostaríamos, restrições que são violadas frequentemente não devem possuir um peso relativamente alto. Começaremos agora a descrever como modelos de Entropia Máxima (não apenas o caso particular do Maxent PL) determinam matematicamente essas propostas informais.

Chamemos de D um conjunto de dados fonológicos do PB consistindo de *tokens* de *palavras*, como por exemplo, $[pratika, kaza, \varepsilon\text{jkada}, pratika, pr\acute{o}za]$ ¹³. Podemos então definir a conta empírica de violações de uma *restrição* C_i no conjunto D como segue:

¹³ É importante notar que embora estejamos falando de conjuntos, não desejamos que $\{pratika\} = \{pratika, pratika\}$, pois queremos contar os *tokens* de um mesmo tipo e por isso utilizamos uma notação sem as chaves. Assim o leitor deve considerar $[pratika] \neq [pratika, pratika]$ e pensar nesse tipo de notação quando falarmos do conjunto de dados D .

$$Contagem(C_i) = \sum_{d \in D} * C_i(d)$$

E, conseqüentemente, podemos definir uma estatística importante, que é a média empírica, ou *valor esperado empírico* de C_i como segue:

$$E^*(C_i) = \frac{1}{|D|} \sum_{d \in D} * C_i(d)$$

No conjunto de dados que utilizamos como exemplo acima, os resultados da *contagem empírica* e *expectativa empírica* para $*pr$ seriam:

$$Contagem(*pr) = \sum_{d \in D} *pr(d) = 3$$

$$E^*(pr) = \frac{1}{|D|} \sum_{d \in D} *pr(d) = \frac{3}{5}$$

Estas são, como queríamos, a conta empírica do número de vezes que a restrição $*pr$ foi violada em nos dados [pratika, kaza, eřkada, pratika, proza] e a razão desse número em relação ao tamanho da amostra.

Definimos agora o *valor esperado do modelo*, que corresponde à esperança do número de vezes que uma restrição C_i seja violada em Ω :

$$E(C_i) = \sum_{x \in \Omega} P(x) C_i(x)$$

Estamos agora aptos a expressar os dois princípios que configuram um modelo de Entropia Máxima¹⁴:

1. **Regra da Delimitação:** dentro do conjunto S de todos os possíveis modelos P , escolha um P que satisfaça:

$$E^*(C_i) = E(C_i), \text{ para toda } C_i$$

2. **Princípio de Entropia Máxima:** dentro do subconjunto de S determinado pela Regra da Delimitação, escolha o p^* que satisfaça:

$$p^* = \max H(P)$$

onde

H é a *Entropia* dada por:

$$H(P) = - \sum_{x \in \Omega} P(x) \log(P(x))$$

A Regra de Delimitação requer que um conjunto de estatísticas relacionadas a uma categoria de dados seja incorporado no modelo probabilístico que assumidamente gerou a amostra de dados. Como discute Uffink (1995, p.5) a respeito dessa regra, “a simplistic identification of frequencies and probabilities are not tenable in any usual view of the meaning of probability whatsoever, at least not for finite N ”¹⁵. De fato, a motivação para atribuir tamanha importância a certas categorias de realizações não é determinada matematicamente a partir da teoria da probabilidade, mas parece estar atrelada à filosofia de que certas categorias subjacentes aos processos (dos mais variados tipos) exercem forte influência sobre os mesmos. Essa é uma ideia central nos estudos em fonotática.

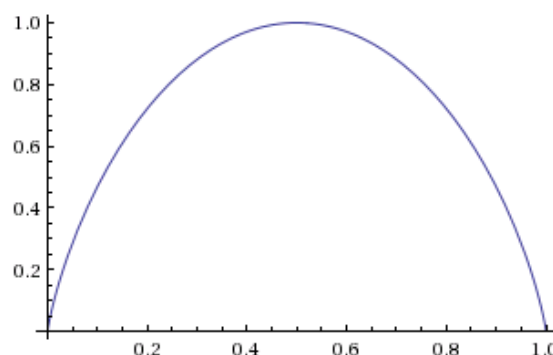
O Princípio de Entropia Máxima, por outro lado, parece balancear a forte aposta dos modelos de Entropia Máxima nas categorias subjacentes aos dados, pois determina

¹⁴ Pedimos ao leitor que lembre de generalizar das restrições para as *features*. Os princípios podem ser lidos em seu formato geral em Berger, Pietra, e Pietra (1996) que utiliza uma definição geral de *feature* como uma função indicadora (o que na verdade é uma simplificação, pois a *feature* em si é uma propriedade dos dados que é mapeada para 0 ou 1 dependendo de sua presença nos dados)

¹⁵ N se refere ao número de observações em um experimento.

que dentre os modelos possíveis que obedecem à Regra de Delimitação, deve-se escolher o de maior aleatoriedade, sem assumir qualquer coisa desconhecida. O conceito de *Entropia* que definimos se originou na Teoria Matemática da Comunicação¹⁶ (SHANNON, 1948), e, podemos dizer que ele é capaz de quantificar o “grau de aleatoriedade” de um modelo probabilístico. Para tomar um objeto simples como exemplo, pensemos no experimento de atirar uma moeda e verificar uma face, tendo como espaço amostral {CARA, COROA}. No caso de uma moeda viciada, digamos, com $P(\text{CARA}) = 0.8$, e $P(\text{COROA}) = 0.2$, os resultados dos experimentos não são tão “supreendentes” quanto uma moeda com $P(\text{CARA}) = 0.5$, e $P(\text{COROA}) = 0.5$, logo dizemos que jogar uma moeda viciada não oferece tanta informação quanto uma moeda justa. Podemos ver no gráfico abaixo como o valor de *Entropia* codifica essas intuições:

Gráfico 01 - função $H(x) = -[x \log_2(x) + (1 - x) \log_2(1 - x)]$, com $x \in [0, 1]$



Com a probabilidade fixada em 1, a função expressa naturalmente os valores de *Entropia* (eixo y) para todas as possíveis distribuições de P com um espaço amostral de dois elementos (eixo x). Por Cálculo, sabemos facilmente que a *Entropia Máxima* se dá no ponto $H(0.5)$, onde a *Entropia* chega a 1 e onde a variável está distribuída o mais uniformemente possível. Como utilizamos o logaritmo com base 2 ao invés do log natural, podemos dizer que naquele ponto é necessário no mínimo 1 bit de informação para codificar o resultado do experimento. Este caso simples, como veremos mais adiante, apresenta resultados que podem ser generalizados para a *Entropia* de qualquer sistema, tal como a propriedade de ser uma função convexa.

Neste exemplo, não há aplicação da Regra de Delimitação porque não usamos dado algum, então o Princípio de Entropia Máxima determina que não assumamos mais

¹⁶ A teoria estabelecida por Shannon na referida obra ficou mais conhecida como Teoria da Informação.

do que sabemos sobre o sistema, escolhendo a distribuição mais uniforme, que corresponde àquela de *Entropia Máxima*. Se tivéssemos um espaço amostral com 3 possíveis elementos $\{x_1, x_2, x_3\}$ conjuntamente a dados que mostrassem que a probabilidade empírica de x_1 é igual a 0.5, e quiséssemos delimitar que nosso modelo tenha $P(x_1) = 0.5$, o Princípio de Entropia Máxima logo determinaria que $P(x_2) = 0.25$ e $P(x_3) = 0.25$, assumindo uma certa informação contida nos dados e considerada (por algum motivo) importante, e nada mais. Como colocam Berger, Pietra, e Pietra (1996, pp.3-4), “Intuitively, the principle is simple: model all that is known and assume nothing about that which is unknown. In other words, given a collection of facts, choose a model which is consistent with the facts, but otherwise as uniform as possible”.

Infelizmente, as soluções analíticas que acabamos de apresentar para encontrar p^* só se aplicam em casos específicos e extremamente simples como nossos exemplos, mas a solução geral do problema não pode ser escrita explicitamente, fazendo com que a utilização de métodos numéricos de otimização sejam utilizados (BERGER et. al., 1996, p.8). Em nosso interesse particular, a *fonotática*, por exemplo, estamos diante de um espaço amostral infinito e, como é de se esperar, um conjunto com pelo menos algumas dezenas de *restrições*. O resultado mais importante para se ter em mente, entretanto, é de que, independentemente do método utilizado para encontrar p^* , p^* é sempre bem-definido e sempre unicamente determinado. (BERGER et al., 1996, p.8).

1.1.2 Cálculo dos pesos

Descreveremos agora os métodos utilizados pelo Maxent PL para encontrar seus pesos uma vez que já se tem um conjunto de dados e um conjunto de restrições. Hayes e Wilson (2008, p.386) partem de um teorema demonstrado por Pietra, Pietra, e Lafferty (1997) (embora já assumido em Berger, Pietra, Pietra (1996)), o qual afirma que se definirmos $P(x)$ como fizemos em 1.1, e aplicarmos a Regra da Delimitação para um dado conjunto de restrições com base em conjunto de dados D , então encontrar p^* é equivalente a encontrar o modelo que maximiza $P(D)$.¹⁷

Assumindo que os elementos $d \in D$ são independentes¹⁸, logo a *probabilidade de D* é dada por:

¹⁷ Passamos a compreender, assim, porque $P(x)$ é definido daquela maneira.

¹⁸ Sabemos que isso pode parecer absurdo se considerarmos se o conjunto de dados corresponde a um texto, pois uma *palavra* condicionaria a próxima devido a fatores sintáticos e semânticos; porém, a partir

$$P(D) = \prod_{d \in D} P(d)$$

Utilizando o fato que funções logarítmicas são monótonas estritamente crescentes, o Maxent PL, por eficiência computacional, maximiza $\log(P(D))$ ao invés de $P(D)$. A busca começa com os pesos w_i recebendo o valor 1, e faz uma busca iterativa ascendente pelo conjunto de pesos que maximiza $\log(P(D))$. Obviamente, o algoritmo utilizado não calcula $\log(P(D))$ para todos os pontos no espaço multidimensional (já assumindo que estamos lidando com dezenas de pesos e, conseqüentemente, dimensões). Ao invés disso, inicia-se no ponto dado quando todos os pesos estão fixados em 1 uma busca iterativa ascendente guiada pelo vetor gradiente local em cada ponto.

O vetor gradiente é formado pelas derivadas parciais de $\log(P(D))$ em relação à cada w_i . A busca por p^* termina heurísticamente quando a distância entre o vetor gradiente e o vetor nulo é um número arbitrariamente próximo a zero (os autores não dizem qual foi especificamente utilizado). É importante notar que só é possível utilizar este método porque sabemos que a função $\log x$ é côncava, e, portanto, não apresenta máximas locais que poderiam ser confundidas com uma máxima global de maneira terminar o aprendizado prematuramente. Dentre os vários algoritmos possíveis para ascender uma superfície a partir dos gradientes, Hayes e Wilson (2008, p. 387) empregam o algoritmo chamado método do gradiente conjugado (PRESS et al., 1992) para realizar a tarefa no Maxent PL. Ainda precisamos discutir, contudo, uma parte importante da otimização pelo método que acabamos de descrever: como os vetores gradientes são calculados.

Mais duas heurísticas são empregadas para calcular os componentes dos vetores gradientes. De acordo com Hayes e Wilson (2008, p.388), já foi demonstrado por Pietra, Pietra, e Lafferty (1997) que (ver apêndice A para um comentário importante sobre tal resultado):

$$\frac{\partial}{\partial w_i} \log(P(D)) = \text{contagem}(C_i) - |D|E(C_i)$$

de agora, o leitor já pode pensar nos dados como uma lista de *palavras* retiradas aleatoriamente de um corpus (e.g. dicionário, conjunto de jornais).

Entretanto, não é possível calcular $E(C_i)$ exatamente uma vez que nosso espaço amostral é infinito¹⁹. Para obter uma solução aproximada, Hayes e Wilson (2008, p.389) propõem tratar o espaço amostral como sendo

$$\Omega = \bigcup_{i=0}^m A^i$$

onde

A^i é um conjunto de *palavras* de tamanho i sobre um *inventário fonológico* A , e

m é o tamanho da maior *palavra* encontrada nos dados utilizados para aprendizado;

Ao fazer isso, os autores conseguem evitar o problema da cardinalidade infinita de Ω de uma maneira eficiente e convincente, uma vez que parece haver uma forte restrição nas línguas naturais a *palavras* “muito grandes”.

Contudo, mesmo lidando agora com um conjunto finito, sabemos que sua cardinalidade ainda é muito alta, o que significa fazer uso de algoritmos de alta complexidade computacional para contar o número de violações de cada restrição no espaço amostral. Para diminuir a complexidade, os autores tomam emprestados métodos da Teoria da Otimalidade Computacional, da teoria das Línguas Formais e da Linguística Computacional (ver HAYES&WILSON, 2008, p.389). O método parte da premissa de que é possível conhecer certas propriedades de uma *língua* ao representá-la como uma máquina de estados finita (ver CHISWELL, 2009, p.5 para definição matemática destas máquinas).

Como os autores não chegam a descrever matematicamente o método (apenas indicam as referências de onde tomaram emprestados os conceitos e as técnicas necessárias) e essa questão diz respeito apenas à implementação (e.g. busca pela diminuição de complexidade computacional), não possuímos informação ou motivação

¹⁹ De fato, pela definição de *restrição* que utilizamos acima, toda restrição é violada um número infinito de vezes em nosso espaço amostral. Portanto, para poder trabalhar com um espaço amostral infinito e entender em que tipos de propriedades estamos interessados, devemos pensar em um subconjunto finito do espaço amostral, como veremos pouco mais abaixo. Para ajudar na intuição, basta pensar que a soma da *boa-formação* do complemento de nosso subconjunto finito (i.e. de cada um dos infinitos elementos que restou no espaço amostral é um número infinitesimal e de pouca relevância).

suficiente para tentar descrever em detalhes o método. Ficaremos satisfeitos, portanto, em compreender a definição $E(C_i)$, seu papel no modelo, e saber que seu valor é aproximado (embora não entendamos precisamente como os autores implementaram o cálculo em computador).

1.2 As restrições

Como já deve ter sido notado, os princípios que caracterizam um modelo como sendo de Entropia Máxima não determinam o conjunto de restrições a ser utilizado, e sim qual distribuição P escolher quando já se tem definido um conjunto de restrições e uma amostra de dados. Assim, o modelo que caracterizamos até então poderia ser qualquer modelo fonotático de Entropia Máxima²⁰, e não especificamente o Maxent PL de Hayes e Wilson (2008). O que diferencia os infinitos possíveis modelos fonotáticos de Entropia Máxima, portanto, é a maneira como o conjunto de restrições é construído. Nada impede, por exemplo, que o linguista determine esse conjunto *a priori*, o que é a prática mais comum em fonologia. A proposta do Maxent PL, por outro lado, é aprender as restrições automaticamente a partir de alguns dados de entrada e um algoritmo de aprendizado bem definido.

É importante destacar, antes de prosseguirmos, que o Maxent PL constitui “uma proposta teórica sobre aprendizagem de língua.” (HAYES&WILSON, 2008, p.390). Os autores iniciam postulando que o aprendiz possui uma Gramática Universal (doravante GU) consideravelmente restrita, que determina apenas um conjunto de traços distintivos (CHOSMKEY&HALLE, 1968) (embora nenhum conjunto específico seja determinado pelos autores) e os formatos que as restrições devem ter.

No quadro abaixo, apresentamos os traços distintivos que utilizaremos nos exemplos e no restante da dissertação (i.e. em nossa aplicação do Maxent PL ao PB.) Escolhemos apenas um número de traços suficiente, como veremos, para representar unicamente cada *fonema* do *inventário fonológico* que definimos para o PB. Fora o cumprimento desta condição, não há nada de especial em relação a esse conjunto específico de traços além do fato de já serem bem estabelecidos e utilizados na literatura fonológica (CLEMENTS&HUME, 1995; NATHAN, 2008):

²⁰ Ou de qualquer outra natureza se substituíssemos nossas restrições (objetos especificamente fonotáticos) pelo conceito mais geral de *feature* que discutimos anteriormente.

Quadro 01 – Conjunto de traços distintivos postulados para GU²¹

Tipo geral	Traços
Classes principais	<ul style="list-style-type: none"> • Consonantal (cons) • Silábico (syll) • Sonoro (son) • Aproximante (appr)
Cavidade	<ul style="list-style-type: none"> • Coronal (cor) • Dorsal (dor) • Labial (lab) • Alto (high) • Baixo (low) • Recuado (back) • Anterior (ant)
Maneira de articulação	<ul style="list-style-type: none"> • Contínuo (cont) • Nasal (nasal) • Lateral (lat)
Fonação	<ul style="list-style-type: none"> • Vozeado (voz) • Glótis difusa (spread)

Em nossos exemplos da subseção anterior, tais como **pr* e **pf*, as restrições eram representadas a partir de seus elementos atômicos, os *fonemas*. Entretanto, o Maxent PL, na verdade, utiliza traços distintivos (i.e. elementos subatômicos dos *fonemas*) para representar suas *restrições*²². Para ilustrar, consideremos um *inventário fonológico* contendo {p, t, k, b, d, g}, e vejamos uma restrição que utiliza traços distintivos e que já está em um dos formatos que postularemos, em seguida, ser determinado pela GU:

²¹ Alguns traços como ‘lateral’ e ‘nasal’ foram categorizados no SPE como traços de cavidade, e muitos autores categorizam os traços ‘alto’, ‘baixo’, ‘anterior’ e ‘recuado’ como traços de vogais, diferentemente do SPE, por exemplo, que os tratam simplesmente como traços de cavidade. Em qualquer caso, estas categorias servem apenas para sinalizar o tipo de informação fonética que os traços pertencentes a elas carregam, mas por si só não possuem qualquer utilidade no formalismo de teorias fonológicas que não utilizem geometria de traços (CLEMENTS, 1985). Portanto, nossa categorização foi puramente didática, deixando o leitor livre para categorizar da maneira que preferir em sua mente já que isto não afetará qualquer resultado de nosso trabalho.

²² Assim como os *fonemas*, trataremos os traços distintivos aqui como primitivos.

$$* \begin{bmatrix} -cont \\ -soan \\ -solt \end{bmatrix} \begin{bmatrix} -cont \\ -soan \\ -solt \\ +voz \end{bmatrix}$$

Como já é sabido em Fonologia Linear, cada matriz de traços é associada ao conjunto de *fonemas* (i.e. classe natural) que possui as especificações dos traços contidos na matriz. No exemplo acima, a primeira matriz está associada ao conjunto de oclusivas {p, t, k, b, d, g}, e a segunda matriz está associada às oclusivas que são vozeadas, ou seja, {b, d, g}; portanto, sendo (x_1, x_2) qualquer elemento do conjunto I^2 (todas as *palavras* de tamanho 2) sobre o *inventário fonológico* I que contém as oclusivas, a representação acima nada mais é do que um conjunto de *restrições* R_1 definido por:

$$R_1 = \{w \in R | w \in I^2 \wedge x_1 \in \{p, t, k, b, d, g\} \wedge x_2 \in \{b, d, g\}\}$$

Portanto, pela nossa definição de *restrição* como *palavra*, o exemplo que demos acima é, na verdade, um conjunto de *restrições* e não uma única *restrição*. A economia formal dessas representações é semelhante ao uso de matrizes de traços em *regras de reescrita*. Em nosso caso, por exemplo, R_1 abarca todas as *restrições* a sequências de duas oclusivas nas quais a segunda é vozeada, o que equivale a 18 *restrições*. Entretanto, é importante lembrar também que cada *fonema* em um *inventário fonológico* deve ter uma representação única em matriz de traços, o que significa que uma única *restrição* pode ser representada também com matrizes de traços, embora na prática em fonologia, estaremos lidando na grande maioria dos casos com conjuntos de *restrições* como R_1 , e não com *restrições* individuais como $*pr$. Portanto, por abuso de terminologia, chamaremos essas representações com traços distintivos também de *restrições* durante o resto da dissertação, acreditando que o contexto tornará claro o significado que estamos empregando em cada caso.

A GU determina dois possíveis formatos para as *restrições*. Acabamos de ver um exemplo específico de um deles, e agora apresentamos sua forma geral (HAYES&WILSON, 2008, p.391):

$$* \begin{bmatrix} \alpha F \\ \beta G \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \begin{bmatrix} \gamma H \\ \delta I \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \dots \begin{bmatrix} \varepsilon J \\ \zeta H \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

onde:

F, G, \dots são traços distintivos,

α, β, \dots são mapeados para um elemento de $\{+, -\}$

Não há limitações definidas para o número máximo de matrizes que a *restrição* pode utilizar, embora, como discutiremos mais adiante, raramente precisaremos de mais de 4 matrizes para representar as restrições mais pesadas das línguas naturais. O número máximo de colunas de uma matriz é, como podemos ver, 2, e, o número máximo de linhas, é, obviamente, igual à cardinalidade do conjunto de traços distintivos postulado.

Sejam m_1, m_2, \dots, m_n matrizes como as que acabamos de determinar para o primeiro formato das *restrições*. Assim, se as matrizes m_1, m_2, \dots, m_n estão, cada uma, associadas respectivamente aos conjuntos de *fonemas* f_1, f_2, \dots, f_n , a forma geral de uma *restrição* R envolvendo as matrizes pode ser definida formalmente como:

$$*m_1, m_2, \dots, m_n = R = \{w \in R \mid w \in I^n \wedge x_1 \in f_1 \wedge x_2 \in f_2, \dots, x_n \in f_n\}$$

Mostramos agora o segundo e último formato possível que as *restrições* podem tomar:

$$* \begin{bmatrix} {}^\wedge \alpha F \\ \beta G \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \begin{bmatrix} \gamma H \\ \delta I \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \dots \begin{bmatrix} \varepsilon J \\ \zeta H \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

A única diferença em relação ao primeiro formato foi a adição do chamado operador de complementação ‘ $^\wedge$ ’ a uma e apenas uma das matrizes. Colocamos ‘ $^\wedge$ ’ na primeira matriz, mas poderia estar em qualquer uma delas (contanto que houvesse apenas

um). O operador é utilizado para formalizar uma *restrição* com implicação lógica (HALLE, 1959; STANLEY, 1967; CHOMSKY&HALLE, 1968; PRINCE&SMOLENSKY, 1993; apud HAYES&WILSON, 2008, p.391). A presença do operador em uma matriz indica que aquela matriz representa todos os *fonemas* (de um dado *inventário fonológico*) que não estão associados à classe natural representada pela matriz. A presença do operador ‘^’ restringe tudo que não seja a classe natural representada pela matriz que o contenha, daí a ideia de “implicação lógica”; ou seja, a classe natural associada àquela matriz é condição necessária para que não haja violação de restrição. Tomemos um exemplo para o PB:

$$* [+cons] \left[\begin{smallmatrix} ^+ lat \\ +ant \end{smallmatrix} \right] [+sil]$$

Esta *restrição* é violada por qualquer *palavra* do PB que não contenha um /l/ entre uma consoante e uma vogal (i.e. no contexto $/[+cons] ___ [+sil]$). Sendo I novamente nosso *inventário fonológico*, *consoante* o conjunto de todas as consoantes em I , e *vogais* o conjunto de todas as vogais em I , podemos, então, definir matematicamente esta *restrição* com implicação da seguinte maneira²³:

$$R_{i1} = \{w \in R | w \in I^3 \wedge x_1 \in \text{consoante} \wedge x_2 \in I \wedge x_2 \notin \{l\} \wedge x_3 \in \text{vogais}\}$$

Note que ‘ $x_2 \in I \wedge x_2 \notin \{l\}$ ’ é simplesmente o complemento de $\{l\}$ em relação a I , denotado por $I \setminus \{l\}$. Logo, supondo que temos as m_1, m_2, \dots, m_n associadas a suas respectivas classes naturais f_1, f_2, \dots, f_n , podemos definir o caso geral de *restrições com implicação* dada por $* ^m m_1, m_2, \dots, m_n$ (com o operador na primeira matriz):

$$* ^m m_1, m_2, \dots, m_n = R_i = \{w \in R_i | w \in I^n \wedge x_1 \in I \setminus f_1 \wedge x_2 \in f_2, \dots, x_n \in f_n\}$$

Do ponto de vista formal, em comparação com o primeiro formato de *restrição* que apresentamos, tudo que precisamos fazer é substituir $x_i \in f_i$ por $x_i \in I \setminus f_i$, onde i é o index da matriz que contém o operador de complemento ‘^’. Entretanto, este segundo

²³ Não confundir o operador de complementação “^” com o símbolo lógico de conjunção “^” que utilizaremos na definição do conjunto.

formato leva a uma economia representacional um pouco maior e adiciona uma espécie de “efeito implicacional” às *restrições*.

Além dos traços distintivos e dos formatos de restrições determinados pela GU, é postulado que o aprendiz já possui também o *inventário fonológico* da língua-alvo ao seu dispor no início do aprendizado das *restrições*. Obviamente o *inventário* não é parte da GU; portanto, o Maxent PL não oferece uma proposta teórica de como um *inventário* é construído pelo aprendiz a partir dos dados do ambiente, partindo então de uma etapa posterior na qual se postula que o aprendiz já possui um *inventário fonológico*. Por fim, assim como no aprendizado dos pesos, o aprendiz tem acesso a um conjunto de dados fonológicos. Agora que já conhecemos todos os dados iniciais que o Maxent PL possui, podemos, finalmente, descrever como o conjunto de *restrições* é construído pelo modelo a partir dessas informações.

1.2.1 Espaço total de restrições

Dados um conjunto de traços distintivos e um *inventário*, chamemos de $|N|$ o número máximo de classes naturais que podem ser definidas a partir dos traços. Sabendo da existência desse número, sabemos também, por combinatória básica, que o número máximo de *restrições* que podemos construir é dado por:

$$\sum_{i=1}^n |N|^i$$

onde

n é o número máximo de matrizes que as *restrições* podem utilizar.

Entretanto, como temos dois possíveis formatos para as matrizes, devemos considerar o número de classes naturais em matrizes sem implicação lógica (chamemos de $|N_1|$) e com implicação lógica (chamemos de $|N_2|$), e alterar a fórmula para conter as possíveis combinações desses dois tipos:

$$\sum_{i=1}^n |N_1|^i + i(|N_2| \cdot |N_1|^{i-1})$$

Como dissemos antes, não há um número determinado para n ; porém, desconhecemos qualquer língua natural que apresente uma “forte” restrição (i.e. violada com baixa frequência) a sequências de mais de 4 *fonemas* ou classes de *fonemas*. Nas aplicações do Maxent PL realizadas por Hayes e Wilson (2008), por exemplo, n foi fixado em 3 para o inglês, e 4 para as línguas shona e wargamay. O primeiro termo do somatório inclui o número de todas as possibilidades de restrições sem implicação lógica, e o segundo termo inclui o número de possibilidades com implicação lógica contendo o operador ‘^’, como definimos, em apenas uma matriz.

Esta fórmula expressa, portanto, o tamanho do espaço total de possíveis restrições que o Maxent PL pode aprender, e, uma solução simples seria tomar o espaço inteiro como constituindo o conjunto de *restrições*, evitando o problema de seleção. Entretanto, vejamos no quadro como o tamanho do espaço cresce para valores de n entre 1 e 4, assumindo que $|N_1| = 97$ e $|N_2| = 90$ (no apêndice 02, deixamos um pequeno programa escrito em Python que utilizamos para calcular os valores abaixo e que pode ser reutilizado no cálculo do espaço total para outros valores de N_1 e N_2):

Tabela 01 – Espaço total de restrições em relação a n , para $|N_1| = 97$ e $|N_2| = 90$

N	Tamanho total do espaço
1	187
2	27.056
3	3.480.159
4	420.571.720

Os valores $|N_1| = 97$ e $|N_2| = 90$ são na verdade os números encontrados por Hayes e Wilson na aplicação do inglês, e, acreditamos, que no PB teremos números próximos a estes. Os autores, como dissemos antes, trabalharam com $n = 3$ para o inglês, o que já oferece um número exorbitante de possíveis *restrições*. Contudo, na referida aplicação, os autores estavam interessados em capturar padrões fonotáticos relacionados aos onsets silábicos do inglês; em nosso caso, estaremos interessados em *restrições* no domínio da sílaba; logo, sabendo que a maior sílaba do PB é de tamanho 4, nosso n precisa ser 4, como foi o caso para shona e wargamay (ver apêndice B para acessar a definição de ‘sílabas do PB’ e ‘onset’ que estamos adotando ao fazer uso dos termos nesta dissertação). Neste caso, ampliamos ainda em milhões o tamanho do espaço de *restrições*.

De qualquer maneira, o ponto crucial aqui é que precisamos de um mecanismo de aprendizado que (HAYES&WILSON, 2008, p. 393):

Given a large set of possible constraints as just defined, we must next form them into a grammar. Since, as already noted, we cannot simply weight all possible constraints, our learner must be made more discerning: it needs a way to home in early on the constraints that are important for characterizing the target language.

Para concluir nossa descrição do Maxent PL, portanto, veremos então como o modelo realiza essa tarefa de seleção e aprende um subconjunto do espaço definido acima para compor a *gramática fonotática* de uma dada *língua*.

1.2.2 Seleção de restrições: acurácia e generalidade

O aprendizado é guiado simplesmente por duas propriedades das restrições (“acurácia” e “generalidade”) definidas pelos autores. Devemos começar pela *acurácia*, denotada por A , de uma *restrição* C_i , que é definida com valores que já conhecemos:

$$A(C_i) = \frac{\text{Contagem}(C_i)}{E(C_i)|D|}$$

onde

$|D|$ é a cardinalidade do conjunto de dados

No artigo original, a acurácia está definida como $\text{Contagem}(C_i)/E(C_i)$ ($O(C_i)/E(C_i)$, na notação de Hayes e Wilson (2008)). Entretanto, como vimos no apêndice 01 em relação ao gradiente (1.1.2) utilizado para o aprendizado dos pesos, esta é uma simplificação que os autores fizeram durante a descrição do Maxent PL, enquanto, na prática, $E(C_i)$ é sempre escalonado por $|D|$,

Para utilizar esta propriedade na seleção de restrições, o Maxent PL determina então uma escala de intervalos de acurácia para ser utilizada iterativamente. Para ilustrar, tomemos a seguinte escala com quatro intervalos:

$$A(C_i) < 0.01, A(C_i) < 0.1, A(C_i) < 0.2, A(C_i) < 0.5$$

O algoritmo do Maxent PL primeiramente seleciona, do espaço total, as restrições que satisfazem $A(C_i) \leq 0.01$, em seguida $A(C_i) \leq 0.1$, e assim por diante, até selecionar todas as restrições que satisfazem o último intervalo da escala (em nosso exemplo, $acurácia \leq 0.5$). Contudo, estamos fazendo uma simplificação aqui. O leitor já deve ter percebido que quanto maior a distância entre o numerador da acurácia ($Contagem(C_i)$) e o seu denominador ($E(C_i)|D|$), (com $Contagem(C_i) \leq E(C_i)|D|$), menor será a acurácia da restrição. Portanto, se uma restrição é violada poucas vezes nos dados da língua mas já era esperado que isso ocorresse (i.e. o valor de $E(C_i)$ é relativamente próximo ao de $Contagem(C_i)$), a violação desta restrição não deveria provocar a mesma intuição de boa-formação que uma restrição cuja distância entre $E(C_i)$ e $Contagem(C_i)$ seja maior. Entretanto, se utilizarmos diretamente o valor de $A(C_i)$, nos deparamos com o seguinte caso problemático:

$$\frac{Contagem(C_i)}{E(C_i)|D|} = \frac{0}{10} = \frac{0}{100} = \frac{0}{n} = 0, \text{ onde } n \in \mathbb{N}$$

Como podemos observar, no caso em que $Contagem(C_i) = 0$, a intuição se perde nos números. Com isto em vista, Hayes e Wilson (2008, p.393), ao implementar a acurácia ao Maxent PL, substituem o uso direto de $A(C_i)$ por um limite superior de confiança estatística para esse valor²⁴. Para o cálculo do limite, a escolha dos autores foi pelo nível de significância $\alpha = 0.975$, que, de acordo com a experiência deles, “helps exclude pointless constraints from the learned grammars without also excluding constraints with explanatory merit.” (HAYES&WILSON, p.393). Utilizando este método, os autores afirmam que a “difference in accuracy between 0/10 and 0/1,000 comes out not as 0 vs. 0, but as 0.22 vs. 0.002.” (p.393). Por fim, ainda em relação à implementação do aprendizado de *restrições*, $E(C_i)$ não é aproximado utilizando uma máquina de estados finita como é feito para o aprendizado dos pesos (como vimos em 1.2.2):

Instead, we take a large random sample from the set Ω of all possible phonological representations. When the sample is sufficiently large and is drawn according to well-established techniques (Della Pietra, Della Pietra, and Lafferty 1997, MacKay 2003), the average number of violations in the sample provides a fairly accurate estimate of the expected value for Ω as a whole.

²⁴ Por conveniência, continuaremos falando apenas de $A(C_i)$ ao tratar do aprendizado de *restrições*, mas o leitor deve ter em mente que na implementação não se utiliza diretamente seu valor.

A “generalidade” responde à seguinte pergunta: em qual ordem selecionar as restrições que satisfazem o mesmo intervalo de acurácia durante cada iteração da escala? Naturalmente, para cada intervalo I da escala de acurácia haverá um conjunto com mais de uma restrição tal que $A(C_i) \in I$; logo, uma escala de “generalidade” é utilizada para determinar a ordem de seleção das restrições de um mesmo conjunto como este. Para determinar a ordem nessa escala, utilizam-se duas propriedades das restrições: 1) o número de matrizes; 2) número de segmentos aos quais a restrição está associada. Estas propriedades são então utilizadas da seguinte maneira para criar uma ordem de generalidade entre restrições:

1. Uma restrição C_1 é mais geral que outra restrição C_2 se o número de matrizes de C_1 for menor que o número de matrizes de C_2 ;
2. Se o número de matrizes de C_1 for igual ao de C_2 , então a matriz associada a mais segmentos é mais geral;

Não sabemos o que acontece se no segundo caso o número de segmentos associados às restrições também forem iguais. Os autores não informam a este respeito, provavelmente porque esse é um evento muito pouco provável²⁵. Isto posto, o Maxent PL aprende primeiramente as restrições mais gerais, indo então das mais gerais para as menos gerais em cada intervalo de *acurácia*. Sobre a motivação por trás do uso do número de matrizes, os autores explicam o seguinte (HAYES&WILSON, 2008, p.394):

This procedure is effective, because longer sequences can often be assessed on the basis of the shorter sequences they contain. For instance, the wellformedness of a consonant cluster *C1C2C3* is usually determined by that of *C1C2* and *C2C3* (Greenberg 1978, Clements and Keyser 1983, Pierrehumbert 1994). In such cases, early discovery of simple, widely applicable constraints obviates the need for more complex ones.

Em relação à utilização do número de segmentos associados às restrições, argumentam também que (HAYES&WILSON, 2008, p.394):

²⁵ De qualquer maneira o algoritmo deve tomar alguma decisão, que provavelmente é a seleção da restrição que foi avaliada primeiro (i.e. a restrição cujas propriedades ‘número de matrizes’ e ‘segmentos associados’ foram calculadas primeiro).

From the same principle it follows that among constraints of equal length, one should first search those whose matrices contain the most general featural expressions. The classic way of assessing featural generality is the feature-counting metric of *SPE*. However, in keeping with our overall emphasis on natural classes instead of their featural expressions, we suggest that the value of a constraint is proportional to the number of segments contained in its classes, and our metric sorts constraints of a given length on this basis.

Com estes comentários, encerramos a descrição do Maxent PL. Agora sabemos como a *fonotática* é definida neste modelo e como seus componentes (restrições e pesos) são aprendidos, resultando em uma *gramática fonotática*. Abaixo apresentamos uma versão levemente modificada do Maxent PL, em nossa linguagem algorítmica, baseada em Hayes e Wilson (2008, p. 394). As modificações se dão principalmente nos termos e na notação (utilizamos termos que definimos em nossa descrição do modelo neste capítulo), com a exceção de um dos dados de entrada, o qual retiramos por não ser input obrigatório e porque só escreveremos a seu respeito em 2.1.

Quadro 02 - Maxent PL em Linguagem Algorítmica

Input: um inventário fonológico *I* representado por um conjunto de traços distintivos, um conjunto *D* de palavras da língua *I**, uma escala ascendente de intervalos de acurácia α

1 inicie uma gramática fonotática vazia *G*

2 para cada intervalo de acurácia em α

3 faça

4 selecione a restrição mais geral com acurácia no intervalo (se houver) e adicione-a à *G*

5 calcule os pesos de cada restrição em *G*

6 enquanto uma restrição é selecionada no passo 4

Output: *G*

CAPÍTULO II – Metodologia

2. Objetivos metodológicos

Como vimos na introdução, nosso objetivo é selecionar um conjunto de *palavras* do PB (chamamos de *S*) e medir o nível de correlação Pearson entre a *fonotática* atribuída a cada uma das formas em *S* por uma *gramática (fonotática)* do PB aprendida pelo Maxent PL (chamamos de vetor *M*) e a *fonotática* atribuída a estas mesmas formas por falantes nativos do PB através de um experimento (chamamos de vetor *E*). Em 2.1, descrevemos a metodologia utilizada para aplicar o Maxent PL ao PB e, assim, obter uma *gramática* do PB (mais precisamente, nas subseções 2.1.1 a 2.1.4). Com base nos resultados da aplicação, selecionaremos os elementos de *S* e construiremos *M* (2.1.5). Por fim, em 2.2., descreveremos o experimento montado para construir *E*. Os resultados da correlação serão apresentados no capítulo 3.

2.1 Construindo M

O Maxent PL está disponível em forma de software gratuito (HAYES et al., 2008). Portanto, como já vimos, para aplicar o aprendiz a uma dada *língua*, precisamos fornecer, no mínimo, os seguintes inputs: i) o *inventário fonológico* da *língua-alvo* representado a partir de traços distintivos; ii) uma amostra de dados (i.e. *palavras*) da *língua-alvo*. Ambos os inputs para o programa devem consistir em arquivos no formato de texto (.txt). Já vimos qual o papel destes inputs no capítulo I, e, agora, descreveremos como construímos tal material para o PB.

2.1.1 O inventário fonológico em traços distintivos

Na construção do inventário, embora não seja obrigatório, seguimos as recomendações de Hayes e Wilson para diminuir o espaço total de *restrições*, e utilizamos subespecificação privativa (também chamada de radical) e contrastiva (STERIADE, 1995) para representar o *inventário* do PB. Apresentamos primeiramente o quadro de consoantes:

Quadro 03 – Consoantes do PB em traços distintivos

	<i>p</i>	<i>b</i>	<i>t</i>	<i>d</i>	<i>k</i>	<i>g</i>	<i>f</i>	<i>v</i>	<i>s</i>	<i>z</i>	<i>ʃ</i>	<i>ʒ</i>	<i>h</i>	<i>m</i>	<i>n</i>	<i>ɲ</i>	<i>l</i>	<i>ʎ</i>	<i>r</i>	<i>w</i>	<i>j</i>
<i>Syll</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Cons</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
<i>Appr</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
<i>son</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+
<i>Cont</i>	-	-	-	-	-	-	+	+	+	+	+	+	+	0	0	0	0	0	0	0	0
<i>Nas</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	+	0	0	0	0	0
<i>Voice</i>	-	+	-	+	-	+	-	+	-	+	-	+	0	0	0	0	0	0	0	0	0
<i>Spread</i>	0	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0	0	0	0	0	0
<i>Cor</i>	0	0	+	+	0	0	0	0	+	+	+	+	0	0	+	+	+	+	+	0	0
<i>Ant</i>	0	0	+	+	0	0	0	0	+	+	-	-	0	0	+	-	+	-	-	0	0
<i>Lab</i>	+	+	0	0	0	0	+	+	0	0	0	0	0	+	0	0	0	0	0	+	0
<i>Dor</i>	0	0	0	0	+	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Lat</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	-	0	0
<i>High</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+
<i>Low</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	-
<i>Back</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	-
<i>Round</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	0

No quadro acima, observamos as especificações das consoantes que selecionamos como parte do inventário fonológico do PB associadas a formas de superfície. Só justificaremos as escolhas na subseção seguinte (2.1.2), quando discutirmos a amostra de dados utilizada em nossa aplicação. Os zeros representam a subespecificação dos traços. Na sequência, apresentamos o quadro das vogais e discutimos quais princípios foram utilizados para subespecificar valores de ambos os quadros.

Quadro 04 – Vogais do PB em traços distintivos

	<i>i</i>	<i>e</i>	<i>ɛ</i>	<i>a</i>	<i>o</i>	<i>ɔ</i>	<i>u</i>	<i>ã</i>	<i>ẽ</i>	<i>ĩ</i>	<i>õ</i>	<i>ũ</i>	<i>ə</i>
<i>Syll</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Cons</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Appr</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>son</i>	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>Cont</i>	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Nas</i>	-	-	0	-	-	0	-	+	+	+	+	+	0
<i>Voice</i>	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Spread</i>	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Cor</i>	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Ant</i>	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Lab</i>	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Dor</i>	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Lat</i>	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>high</i>	+	-	-	-	-	-	+	-	-	+	-	+	-
<i>Low</i>	-	-	+	+	-	+	-	+	-	-	-	-	-
<i>back</i>	-	-	-	+	+	+	+	+	-	-	+	+	+
<i>round</i>	0	0	0	-	+	+	+	-	0	0	+	+	-

Não adotamos aqui uma teoria de subespecificação em particular, pois estas além de serem justificadas historicamente com base em *regras de reescrita* (algo que como vimos desde a introdução não nos interessa aqui), também são pensadas para propósitos mais gerais, enquanto teoria cognitiva da representação de traços e com um escopo que vai além da fonologia no nível lexical do tipo que estamos trabalhando aqui. Isso posto, seria desnecessário adicionar ao nosso modelo postulados sobre a subespecificação de traços quando só queremos realizar a tarefa opcional de diminuir o espaço de possíveis *restrições* para o PB. Assim, tratemos as especificações acima como escolhas opcionais baseadas em princípios simples de subespecificação. Como pode ser facilmente

verificado por uma análise dos quadros acima, não especificamos o valor de um traço (i.e. atribuímos zero ao traço) para um *fonema*, quando:

1. Dentro da classe natural determinada pelos valores de *syll*, *cons*, *appr*, *son*, o valor do traço não é necessário para definir o *fonema* unicamente nessa classe;
2. Após atribuir + para *lab*, *cor*, ou *dor*, o traço é do tipo ‘cavidade’ e não é necessário para definir unicamente o *fonema*.

Como exemplo do primeiro princípio, podemos observar que o traço *voice* foi subespecificado na classe das consoantes nasais, pois os traços *cor*, *dor*, e *ant*, são suficientes para distingui-las. Dito de outra maneira, esse traço não é contrastivo na classe de consoantes nasais. Podemos ver o segundo princípio em ação na subespecificação de *lab* e *dor* para o *fonema* /r/ (assim como para todos os outros *fonemas* coronais). Contudo, este fonema, em particular, mostra também que apesar de *cor* já haver recebido o valor +, *ant* e *lat* precisaram ser especificados também para diferenciar o *fonema* de /k/ e /l/, respectivamente.

Concluimos, assim, a metodologia utilizada para construir um dos dados de entrada de nossa aplicação: o *inventário fonológico* definido para o PB representado por traços distintivos. No apêndice D, mostramos o *inventário fonológico* completo no formato que o software do Maxent PL aceita como input. De fato, lá observamos uma imagem do arquivo que utilizamos no programa. Note que os símbolos utilizados lá para representar os fonemas são diferentes daqueles apresentados nos quadros acima. O motivo das alterações será explicado ao final da subseção seguinte (2.1.2).

2.1.2 A amostra de dados

Já sabemos que nosso conjunto de dados deve ser composto por *palavras*, cuja representação equivale às transcrições fonológicas. O que ainda não informamos é: i) o formato em que as transcrições devem ser apresentadas para o software do Maxent PL; ii) o tipo de corpus; iii) o tamanho que a amostra deve possuir. Como foi brevemente sugerido na nota 16, as palavras transcritas devem ser fornecidas ao Maxent PL em forma de lista. Além disso, não se deve utilizar barras ‘/’ ou acento nas transcrições (veremos exemplos mais abaixo). Não há restrição teórica ou metodológica sobre o tipo de corpus

que deva ser utilizado. Entretanto, uma decisão importante que deve ser tomada é se devemos incluir na lista ocorrências (tokens) de *palavras* ou apenas o tipo delas. Para ser mais preciso, se uma *palavra* como /*elégãti*/ ocorre dezessete vezes no corpus, devemos incluí-la dezessete vezes na lista (token) ou apenas uma (tipo)?

Em uma das aplicações do Maxent PL, na qual as duas opções foram comparadas, a utilização de ocorrências produziu resultados de acurácia ligeiramente menores em relação aos dados experimentais. Ademais, os autores acrescentam que (HAYES&WILSON, 2008, p.395):

In general, it appears that the use of type frequencies yields better results in modeling any sort of phonological intuitions based on the lexicon; for discussion, see Bybee 1995, 2001, Pierrehumbert 2001a, Albright 2002b, Albright and Hayes 2003, Hayes and Londe 2006, and Goldwater 2007.

Com base nestes resultados e para manter o parâmetro das aplicações de Hayes e Wilson, decidimos incluir em nossa lista apenas os tipos de palavras encontradas no corpus que selecionaremos. Nosso segundo passo, portanto, é selecionar o corpus que utilizaremos para construir a lista. Como já mencionamos, não há imposição teórica ou metodológica sobre esta decisão. Nossa população é, obviamente, o léxico do PB, que pode ser definido para fins práticos como, por exemplo, todas as *palavras* do PB faladas (expressas foneticamente) ou escritas (expressas graficamente) no Brasil nos últimos 50 anos. Dito isto, sabemos que é impossível realizar uma amostragem probabilística.

Defendemos, contudo, que apesar de não probabilística, seria impossível, no estado atual de conhecimento, saber se estamos fazendo uma amostragem viciada ou não. Isso acontece porque desconhecemos qualquer estudo sobre o PB que informe sobre a relação entre variáveis linguísticas ou extralinguísticas e a frequência de violação de qualquer restrição ou conjunto de restrições fonotáticas, seja em relação ao tipo ou ao número de ocorrência destas. A única maneira de viciar intencionalmente uma amostra, portanto, seria construir um corpus após possuir resultados experimentais. Portanto, qualquer amostra do tamanho mínimo exigido (veremos mais abaixo qual é esse tamanho) e obtida antes de experimentos fonotáticos do tipo que descreveremos em 2.2 poderia servir de input em nossa aplicação.

Entretanto, há algumas opções mais interessantes que outras, seja por estarem “mais próximas” de uma amostragem aleatória, seja por estarem mais próximas de uma situação real de aprendizagem de um sistema fonotático (i.e. utiliza palavras mais

acessíveis no ambiente aos falantes-ouvintes do PB). Um dicionário do PB, por exemplo, não seria uma opção absurda para substituir nossa definição acima de léxico do PB e acabar com o problema da impossibilidade de uma amostragem aleatória simples. Claro que seria uma população muito menor do que a que definimos, além de não conter muito do léxico vernáculo (ou não conter muito da língua escrita, caso o dicionário fosse voltado para o vernáculo); entretanto, acreditamos que qualquer pesquisador concordaria, mesmo que seja mera conjectura, que seria melhor realizar nossa amostragem a partir de um dicionário do que um livro de trava-línguas, no qual muitas restrições que acreditamos não serem violadas com tanta frequência (seja pelo tipo ou ocorrência) em outros tipos de texto são violadas em alta frequência por motivos óbvios. De fato, na aplicação do Maxent PL ao inglês e à língua shona, Hayes e Wilson realizam a amostragem a partir de dicionários.

Já que estamos lidando com um modelo teórico de aprendizagem de língua, outra escolha interessante seria um corpus com palavras frequentemente utilizadas, ou, equivalentemente, facilmente acessíveis a falantes-ouvintes nativos do PB. Talvez o caso concreto mais interessante de um corpus com essa característica fosse um corpus de língua falada ao qual uma criança nativa brasileira teve acesso em seu ambiente. Entretanto, teríamos diversas variáveis a considerar em um caso idiossincrático como este. Para mencionar talvez a mais problemática, o Maxent PL já possui um *inventário fonológico*, mas em qual idade um aprendiz humano adquire o seu? Recairíamos, assim, em um número consideravelmente grande de escolhas arbitrárias que talvez não fosse tão representativa do processo humano de aprendizagem quanto se pudesse pensar inicialmente. Nossa escolha, entretanto, envolve aspectos dos dois exemplos que mencionamos. O corpus escolhido para a transcrição automática consiste de uma lista com as 3000 palavras mais frequentes do jornal ‘Massa!’, que se trata de um “jornal popular baiano do grupo A TARDE com edições diárias na versão on-line e impressa, que tem o objetivo de oferecer notícias e informação de entretenimento e serviço às classes C e D. O novo produto, que é o primeiro jornal popular da Bahia, aposta na linguagem coloquial, mas nem por isso menos informativa.” (FINATTO et al., 2017)

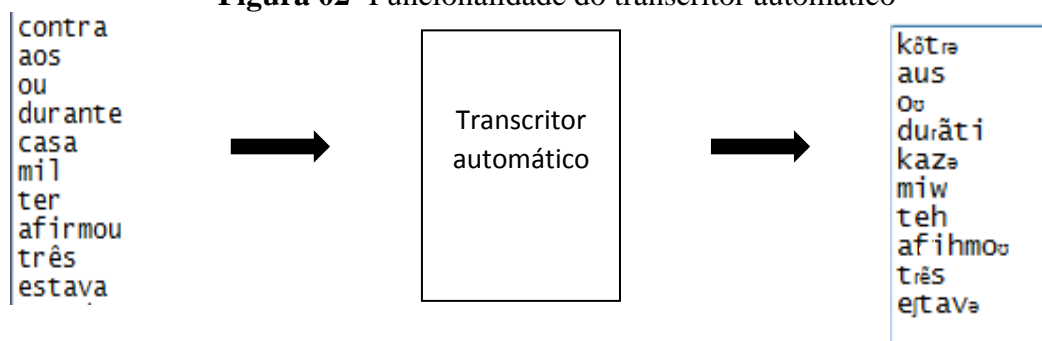
A lista foi construída a partir de uma amostra de 70 textos de notícias publicadas no jornal entre 03 de janeiro e 06 de fevereiro de 2013 e se encontra em domínio público (FINATTO, 2017). Estas palavras poderiam ter sido simplesmente retiradas de um dicionário, mas incluem também o aspecto de serem as mais frequentemente utilizadas em pelo menos uma “modalidade” de texto (i.e. escrito e jornalístico). Por outro lado,

devemos nos indagar se estas variáveis viciaram a amostra. Porém, como discutido antes, não temos informação a este respeito, então tudo que podemos fazer é assumir a conjectura de que duas das variáveis que caracterizam o tipo do corpus (i.e. texto escrito e jornalístico) não são significativas em relação aos tipos de restrições que serão violadas quando comparadas com língua falada e não jornalística²⁶.

A última informação que ainda não descrevemos é o tamanho necessário da amostra. Embora o software funcione com um número menor, no manual do Maxent PL (HAYES&WILSON, 2009, p.8), os autores alertam que “it is wise to make sure that the frequencies total above 3000.” O próprio software também gera uma mensagem de advertência caso a lista utilizada como input não contenha um mínimo de 3000 palavras. Portanto, este é o tamanho mínimo da amostra recomendada pelos autores para obter uma performance satisfatória do aprendiz artificial.

Descreveremos agora como transformamos o corpus escolhido em um arquivo que servisse de input para o Maxent PL. Naturalmente, o corpus escolhido não está transcrito fonologicamente (as palavras estão representadas por grafemas); logo, foi necessário transcrever todas as palavras. Devido ao tamanho da amostra, escrevemos um programa para transcrever automaticamente as representações com grafemas em representações com fonemas²⁷. Nosso programa toma como input uma lista de palavras em representação grafêmica, e devolve a transcrição fonológica das mesmas. Vejamos um esquema:

Figura 02- Funcionalidade do transcritor automático



²⁶ Embora seja interessante observar que não estamos lidando com um texto jornalístico tradicional, o que também adiciona uma mistura de registro que provavelmente quebra com as propriedades que se espera da distinção entre variantes como jornalístico vs. não-jornalístico, ou língua escrita x língua falada. Isto adiciona uma complexidade bem-vinda à amostra.

²⁷ Não parecia haver um corpus do PB que consistisse de um mínimo de 3000 palavras transcritas (ou algo perto disso). Nossas buscas pela internet e o contato com a comunidade de linguistas não retornaram qualquer resultado, o que nos levou a acreditar na inexistência ou indisponibilidade de um corpus dessa natureza. Por este motivo, não selecionamos um corpus já transcrito e escrevemos o programa.

Tendo em vista que a construção de um programa dessa natureza corresponde a um projeto de pesquisa por si só, uma discussão aprofundada de seu algoritmo fugiria de nosso escopo e nos desviaria do objetivo desta dissertação. Portanto, nos ateremos aos aspectos do transcritor automático que são cruciais no desenvolvimento de nossa metodologia. Como estamos lidando com formas de superfície, o primeiro questionamento que deve surgir na cabeça de um fonologista é: quais processos fonológicos²⁸ foram incluídos nas transcrições?

Dentre os diversos possíveis processos descritos na literatura fonológica do PB, escolhemos alguns dos mais gerais e categóricos, ou seja, aqueles que são descritos na maioria das regiões do Brasil sobre as quais conhecemos estudos (CÂMARA JR., 1966; BISOL, 1996; HORA, 1999, 2006; PEREYRON, 2008; HAHN, 2010) e que se aplicam quase sempre:

Quadro 05 – Processos utilizados no transcritor automático

Processo	Exemplo
Vocalização da lateral em coda	natal → nataw
Nasalização das vogais seguidas por nasais em coda	kaNtaR → kãtah
Alçamento da vogal aberta em final de palavra	espada → espadə
Alçamento das vogais médias altas em final de palavra	elefaNte → elefãti
Epêntese em encontros de obstruentes	optaR → opitah
Epêntese em encontros de obstruientes com fricativas	afta → afitə
Assimilação de vozeamento do contexto posterior pelo s	asnu → aʒnu
Vozeamento do s intervocálico	kasa → kazə

Sabemos que alguns itens da lista nem sejam considerados processos por um número de fonologistas (como o segundo ponto, por exemplo), mas esta é dificilmente a questão aqui. Essa decisão teórica pouco importa na presente dissertação, pois já deve estar claro para o leitor que, de acordo com nossa descrição do Maxent PL na introdução e no capítulo I, as relações que postulamos entre formas subjacentes e formas de

²⁸ Utilizamos o termo “processos” aqui em sentido mais geral do que *derivações* de uma forma subjacente a uma forma de superfície por meio de *regras de reescrita*, mas na existência de uma relação entre formas nos dois níveis. Por exemplo, o processo de vocalização da lateral em coda é pensada como uma relação entre / em coda na forma subjacente e w na forma de superfície.

superfície não possui qualquer efeito prático em nossa aplicação. Portanto, embora de um ponto de vista mais geral da teoria fonológica, essa seja uma questão de grande importância, não nos cabe discuti-la aqui. Isso posto, pode-se olhar para o que chamamos de processo acima como uma maneira simples e concisa de informar o leitor, a partir de termos cujos significados são bem conhecidos na área, quais formas de superfície foram utilizadas em nosso transcritor.

Retornando à questão principal, isto é, a de quais escolhas fizemos (independentemente de as chamarmos de processo ou não), podemos observar já nos exemplos acima que incluímos dois processos mais específicos (i.e. mais comuns em uma região menor do país):

- Palatalização das fricativas alveolares antes de /t/, /d/, e /n/²⁹ (e.g. asnu → aʒnu);
- Implementação aspirada do /R/ (e.g. e.g. ɔptaR → ɔpitah).

A escolha pela aspirada reflete a forma mais utilizada no nordeste brasileiro (embora seja também bastante utilizada em outras regiões), enquanto que a palatalização no contexto dado é mais específica, estando associada com dialetos do Estado da Paraíba (HORA, 1999). Para evitar processos mais específicos em nossa amostra de dados, poderíamos ter especificado arquifonemas em nosso inventário (mesmo não sendo formas de superfície), mas não vemos problema em nossa escolha; além disso, ainda podemos testar se o Maxent PL aprenderá alguma *restrição* relacionada à palatalização nos contextos que determinamos. Não incluímos qualquer tipo de harmonia vocálica nas transcrições. Por fim, em relação ao caso problemático das vogais médias, o transcritor só utiliza vogais médias baixas quando os grafemas <e> e <o> são acentuados, ou seja, palavras como <janela> e <ela> são transcritas erradas. Este problema diminui a acurácia do transcritor automático, uma questão que trataremos em mais detalhes abaixo.

Uma última informação importante é que nosso transcritor realiza todos os processos descritos de maneira categórica. Utilizamos dessa idealização por dois motivos. Primeiro, se admitíssemos variação, qualquer frequência que escolhêssemos para as variantes seria arbitrária ou muito específica de algum dialeto. E em segundo lugar, como

²⁹ A inclusão do /n/ como contexto para o processo advém de uma descoberta recente nos dados de João Pessoa em pesquisa ainda não publicada. Há pouco tempo, se pensava que apenas /t/ e /d/ condicionavam o processo.

os processos escolhidos são “quase categóricos”, não estamos tão distantes da realidade empírica.

Um segundo questionamento que deve ser feito em relação ao nosso transcritor diz respeito à já referida acurácia. Tendo como parâmetro que pretendíamos transcrever de acordo com os processos descritos acima, medimos a acurácia como a porcentagem de acerto na transcrição das 100 primeiras palavras do corpus escolhido, o que revelou uma acurácia de 97%. O principal problema para o transcritor são as vogais médias. Sem entrar em detalhes muito técnicos, podemos dizer que, grosso modo, para cada grafema do input (i.e. palavras da lista), nosso algoritmo associa um *fonema* ou nada (compondo assim a transcrição iterativamente) com base em três tipos de informação: i) o número de grafemas que seguem o grafema atual da iteração; ii) os grafemas que antecederam ou precedem o grafema atual da iteração (geralmente até no máximo dois elementos antes e/ou depois); iii) aquilo que já foi transcrito até a iteração atual.

Com base nessas informações, qualquer fonologista que trabalha com o PB já deve estar convencido que seria impossível transcrever corretamente a vogal média (nos casos nos quais duas variantes não são permitidas) quando na verdade este é um problema em aberto na área (i.e. ainda não foi solucionado mesmo de posse de todo seu aparato teórico). A partir de exemplos como <janela>, <ela>, <nela>, <panela>, poderíamos decidir transcrever <e> como ϵ se o grafema que segue <e> for <l>. Entretanto, ainda erraríamos em <pela>, <pelo>, <cabelo>. Embora nos pareça que, estatisticamente, o ϵ antes de <l> seja mais comum, isto é apenas uma conjectura, pois não possuímos estudos em larga escala para confirmar tal proposição. Além disso, teríamos que criar conjecturas para todos os possíveis contextos posteriores e/ou anteriores. Tal procedimento seria desnecessário para nossos propósitos, uma vez que esse tipo de erro de transcrição não afeta padrões fonotáticos que não estejam relacionados especificamente com estes pares de sons (e- ϵ , o- o), os quais poderemos facilmente identificar caso alguma *restrição* associadas a eles seja aprendida em nossa aplicação.

Problemas menores envolvem “nasalização” de vogais em estruturas pouco frequentes, tais como “hiato” (e.g. <ainda>). Há, possivelmente, problemas ainda desconhecidos, e que só descobriremos com mais testes. Por hora, julgamos a acurácia satisfatória para nossa aplicação. No apêndice E1, o leitor pode visualizar as 100 primeiras palavras do corpus em grafemas (i.e. aquelas utilizadas no teste de acurácia) e avaliar as transcrições realizadas pelo nosso programa para estas palavras no apêndice E2. As três transcrições erradas estão marcadas com o comentário “##erro” em E2.

Um problema técnico que deixamos por último é o fato de que o software do Maxent PL não aceita caracteres codificados em Unicode, o que significa que não podemos utilizar transcrições contendo todos os símbolos do IPA. Isso posto, é necessário criar ou utilizar outro alfabeto fonético já existente (e.g. ARPAbet) para os *fonemas* de nosso *inventário fonológico* que possa ser processado pelo software. A recomendação encontrada no manual a este respeito é a seguinte:

Phonetic symbols for the program may have more than one character (e.g., you can use “ch” or “CH” for [tʃ]). They may not begin with digits, nor may they contain apostrophes, colons, or the symbol @. There may be other restrictions as yet undiscovered! It is probably wisest to use ordinary letters for your symbols. (HAYES&WILSON, 2009, p.14).

Uma alternativa simples que funcionou para nós foi criar nosso próprio alfabeto contendo apenas caracteres latinos (codificados em CP-1252, nos sistemas Windows). Feito isto, definimos uma bijeção dos símbolos utilizados pelo nosso transcritor automático para os símbolos de nosso novo alfabeto e escrevemos um programa para realizar o mapeamento automaticamente, ou seja, para tomar como input as palavras transcritas pelo nosso transcritor e devolver a lista de palavras transcritas no novo alfabeto (i.e. uma simples conversão). De fato, só precisamos alterar aqueles símbolos que não eram de nosso alfabeto latino usual ou possuíam acentuação. No quadro abaixo mostramos os símbolos antigos (S.A) que foram alterados e os novos símbolos associados a eles (S.N.). No apêndice E3, mostramos novamente as 100 primeiras palavras do corpus, mas agora transcritas com o novo alfabeto (i.e. com as alterações do quadro abaixo), o que significa que aquela foi a versão que de fato utilizamos na aplicação. Por fim, note que nosso inventário fonológico representado em traços distintivos já havia sido apresentado no apêndice D com este novo alfabeto:

Quadro 06 – Alterações no alfabeto fonológico

S.A.	ʃ	ʒ	ɲ	ɬ	ɛ	ɔ	ã	ẽ	ĩ	õ	ũ	ə
S.N.	X	jj	Nh	lh	ee	oo	Na	en	In	On	un	q

Por fim, embora o corpus escolhido possuísse exatamente o número mínimo do tamanho da amostra de que precisávamos, muitas palavras eram abreviações e estrangeirismos que escolhemos (mais por intuição em relação a cada caso do que por uma definição precisa) não representar o léxico do PB. Alguns exemplos incluem *km*, *kg*,

h, on, the, kiss. No total, excluímos 107 *palavras* dessas duas categorias, e, para alcançar novamente as 3000 palavras sem alterar as características do corpus simplesmente duplicamos as 2897 palavras, totalizando 5786 palavras e preservando as proporções entre as *restrições* violadas.

2.1.3 Parâmetros do software

De posse dos dados de entrada, é possível, finalmente, realizar a aplicação do PB. Contudo, é preciso ainda mencionar que o software do Maxent PL oferece alguns parâmetros e funções extras cujo entendimento dos valores utilizados são essenciais para o tipo de aplicação desejada. O leitor que tiver interesse pode ler sobre todos eles no manual que já citamos acima (HAYES&WILSON, 2009). Para nossa aplicação, apenas 3 parâmetros básicos são importantes e já preparamos o leitor para compreendê-los no capítulo 1. O primeiro deles é o número máximo de *restrições* a serem aprendidas pelo Maxent PL (i.e. o tamanho máximo da *gramática fonotática*). Por razões que veremos mais abaixo, foi necessário fixar um número para tal parâmetro.

O segundo é o número máximo de matrizes que as *restrições* podem utilizar, o qual denotamos em 1.1.2 como n e investigamos como o espaço total de *restrições* se relaciona a tal parâmetro. No software do Maxent PL os valores possíveis para n são 1, 2, 3, ou 4, e o padrão, caso o valor não seja especificado pelo usuário é 3. Como já mencionamos rapidamente em 1.1.2, nosso n será 4, pois este é o tamanho máximo das *sílabas do PB*. Nosso objetivo com tal decisão é deixar disponível ao aprendiz todas as possíveis restrições no escopo da sílaba (mas claro, isto não assegura que todas as *restrições* estarão no escopo de uma única sílaba). O terceiro e último parâmetro importante diz respeito ao formato das restrições. Vimos em 1.1.2 que existem dois formatos disponíveis para o Maxent PL postulados como parte da GU; entretanto, é possível realizar aplicações do Maxent PL sem utilizar restrições com operador de complementação (^). Para tornar disponível ao aprendiz este tipo de restrição, é necessário marcar a opção “Allow complement natural classes” na interface do programa. A imagem abaixo mostra a interface do programa com os parâmetros acima configurados para nossa aplicação:

Figura 03 – Interface de usuário do software do Maxent PL

Como pode ser visto na figura, determinamos que a gramática fonotática aprendida pelo Maxent PL tivesse um máximo de 58 restrições. Tal número foi escolhido após a realização de aplicações testes nos anos de 2015 e 2016 e explicamos agora por quê. Se nenhum número máximo de restrições for determinado, o aprendiz encerra quando aprende todas as restrições que satisfazem o último intervalo definido para a escala iterativa de intervalos de acurácia, ou seja, todas as restrições C_i cujo valor de $O[C_i]/E[C_i]$ seja menor ou igual a uma cota inferior que define o último intervalo determinado para a escala de acurácia³⁰ (e.g. que satisfaça $O[C_i]/E[C_i] \leq 0.5$). A este respeito, os autores alertam que:

Powerful, important phonotactic constraints tend to have low O/E values. The program searches for these constraints first. But, unless you specify otherwise, it will continue to search for constraints, even rather nonuseful ones, essentially to the bitter end—i.e. up to the value of 1 for O/E. (HAYES&WILSON, 2009, p.10)

³⁰ Estamos utilizando $O[C_i]/E[C_i]$ por simplificação, pois como já vimos em 1.2, o Maxent PL utiliza um limite superior de confiança no lugar do resultado direto de $O[C_i]/E[C_i]$.

O padrão do software é, portanto, $O[C_i]/E[C_i] \leq 1$. Entretanto, outro valor pode ser definido no parâmetro básico “Maximum OE for constraints (supplements accuracy schedule)”. Observamos, também na prática, que é altamente recomendável que se altere este valor. Em uma de nossas aplicações testes, na qual não delimitamos um máximo de restrições e mantivemos o padrão $O[C_i]/E[C_i] \leq 1$ para acurácia, o Maxent PL permaneceu aprendendo por 15 dias até que o encerramos manualmente. Isto nos deixa com duas opções não só mais viáveis em relação ao tempo gasto na aplicação como também mais interessantes do ponto de vista teórico (pois retornam gramáticas com restrições mais poderosas): delimitar um último intervalo mais baixo para a acurácia ou delimitar o número máximo de restrições.

A maioria das aplicações em Hayes e Wilson (2008) utilizaram $O[C_i]/E[C_i] \leq 0.3$. Embora entendamos por que valores menores que 1 são mais interessantes, não há motivação teórica para a escolha precisa desse valor. Muito provavelmente os autores decidiram utilizá-lo mais frequentemente após testar iterativamente valores cada vez mais baixos em aplicações testes e chegar a um valor que proporciona um tempo de aprendizado mais efetivo sem perder resultados importantes (i.e. restrições poderosas e gerais para modelagem de dados empíricos da língua da simulação). Infelizmente, 0.3 não se mostrou efetivo em relação ao tempo nas nossas aplicações testes. Novamente, o programa passa dias aprendendo sem ao menos ultrapassar $O[C_i]/E[C_i] \leq 0.01$. De fato, diversos testes revelaram que, para “Maximum gram size” configurado em 3 ou 4 (não realizamos simulações com n menor que 3), e com nossos dados de entrada, o tempo de aprendizado por restrição, em minutos, cresce exponencialmente a partir da restrição 59. Não realizamos um controle preciso deste tempo, mas podemos dizer de maneira aproximada que, após aprender 59 restrições nas primeiras 2 a 3 horas (a qual é exibida como 58 na interface do programa porque a contagem é feita a partir de 0), o Maxent PL demora mais de 24 horas para aprender a restrição 60, e períodos ainda mais prolongados para cada restrição subsequente.

Diante do exposto, resolvemos delimitar o número máximo de restrições como sendo 58, tendo como principal motivação tornar aplicações que utilizem nossos dados de entrada e os demais parâmetros mais facilmente replicáveis. Entretanto, como vimos em 1.1.2, os pesos das restrições são recalculados sempre que uma nova restrição é aprendida, e não temos estatísticas suficientes para estimar quão diferente os pesos finais de uma gramática com mais de 58 restrições (e.g. uma gramática que encerrasse por

$O[C_i]/E[C_i] \leq 0.3$) seria do resultado que obtivemos. Esta não é uma questão trivial. Por este motivo e por outro que explicaremos em 2.4, após apresentar o resultado desta aplicação que descrevemos até então, realizamos uma segunda aplicação que utiliza como dados de entrada apenas os onsets das palavras de nosso primeiro corpus (i.e. os onsets das 5786 da lista).

2.1.4 Resultados das aplicações

A gramática fonotática aprendida pelo Maxent PL a partir dos dados de entrada e parâmetros que discutimos nas subseções 2.1, 2.2 e 2.3 pode ser vista no anexo A.1. Para facilitar a leitura das *restrições* disponibilizamos no anexo A2 o arquivo gerado pelo Maxent PL no qual é informado o conjunto de segmentos associado a cada matriz (classe natural) presente na gramática fonotática em A1. Com raríssimas exceções, a exemplo de “*[-continuant][+word_boundary]”, desconhecemos descrições ou propostas, na literatura do PB, da grande maioria das restrições aprendidas pelo Maxent PL. Há pelo menos três razões plausíveis para isto. A primeira é que modelos fonotáticos probabilísticos são pouco utilizados no Brasil, o que significa que, geralmente, apenas restrições categóricas (i.e. que nunca ou “quase nunca” são violadas no léxico do PB) são a fonte única ou primária de interesse dos pesquisadores (o que tem a ver com os postulados na base das teorias utilizadas, como vimos na introdução). Assim, modelos que partem de uma visão gradiente de fonotática tenderão, naturalmente, a possuir um número maior de restrições, já que restrições de pesos variados são de interesse do ponto de vista teórico e experimental. Em segundo lugar, não podemos desconsiderar também o poder computacional de um método de aprendizado automático para avaliar o espaço total de restrições possíveis comparado ao de um linguista humano.

Por último, o Maxent PL não possui, enquanto componente *a priori*, uma noção de estrutura silábica, a qual permitiria limitar a busca por restrições dentro de um escopo estrutural mais restrito (e.g. onset, coda, rima). Tal ausência pode levar a padrões talvez jamais pensados ou descritos antes por fugirem do escopo mais convencional da análise fonotática (i.e. a sílaba e seus componentes), mas que podem se revelar enquanto restrições poderosas. Ao mesmo tempo, a ausência de um componente silábico provavelmente limita a capacidade de apreender padrões altamente poderosos e gerais que são já bem conhecidos, e os quais motivam, historicamente, as propostas de estruturas silábicas. De fato, é esta última conjectura que mais interessa ao nosso problema de

pesquisa. Será que a falta de um componente silábico no modelo resulta em uma gramática fonotática de menor acurácia em relação aos dados experimentais? Para mostrar que esta não é uma questão trivial, abriremos um parêntese nesta subseção para falar da proposta da linha de base indutiva de Hayes e Wilson (2008).

Como já vimos em 1.1.2, o Maxent consiste em uma proposta teórica de aprendizagem de línguas (i.e. candidato à teoria cognitiva), e no estado inicial do modelo, tudo que o aprendiz possui é uma GU contendo apenas um conjunto de traços distintivos e os formatos das restrições (dois tipos). Fora isto, assume-se ainda que o aprendiz possua também um inventário fonológico especificado nos traços providos pela GU. Esta, contudo, não é a história completa. Hayes e Wilson não colocam a GU que acabamos de descrever novamente como a versão definitiva para o modelo. Ao invés disso, é proposto um programa de pesquisa em fonotática que deve fazer uso de três princípios: gradiência, expressividade, linha de base indutiva (HAYES&WILSON, 2008, pp.381-382).

Os autores defendem, em relação à gradiência, que os resultados experimentais que discutimos na introdução oferecem evidência de intuições gradientes de boa-formação, e, que, portanto, modelos de aprendizes fonotáticos devem dar conta dessas intuições. Na introdução, axiomatizamos tal princípio como parte dos fundamentos de modelos fonotáticos probabilísticos, e agora discutiremos os outros dois princípios. A expressividade diz respeito ao rico conjunto de estruturas e fenômenos encontrados nos trabalhos em teoria fonológica nas últimas décadas (p.381), tais como:

[...] long-distance dependencies (e.g., McCarthy 1988), phrasal hierarchies (e.g., Selkirk 1980a), metrical hierarchies (e.g., Liberman and Prince 1977), elaborate interactions with morphology (e.g., Kiparsky 1982), and other areas, each the subject of extensive analysis and research.

Para os autores, portanto, qualquer aprendiz fonotático deverá incorporar resultados de diversas áreas da teoria fonológica, pois os mesmos acreditam que os princípios fonotáticos (i.e. as leis que determinam a boa-formação das formas no âmbito fonológico de uma língua) são multi-classificatórios (tradução de “cross-classifying”), ou seja, fazem uso de diferentes classes a qual uma mesma forma pertence simultaneamente devido à existência de diferentes níveis estruturais (e.g. a posição na sílaba, o nó vocálico adjacente, o segmento adjacente, o tipo de sílaba em uma hierarquia métrica). Entretanto, ao mesmo tempo, os autores propõem o seguinte (HAYES&WILSON, 2008, p.381):

While we have emphasized the primacy of phonological theory, the precise content of the latter remains an area of considerable disagreement. A computational learning model can be used as a tool for evaluating and testing theoretical proposals. The idea is that a very simple theory can provide a sort of *inductive baseline* against which more advanced theories can be compared. If the introduction of a theoretical concept makes possible the learning of phonotactic patterns that are inaccessible to the baseline system, the concept is thereby supported.

Os autores utilizaram na prática a proposta da linha de base indutiva nas aplicações à língua shona. De posse apenas dos componentes de GU que descrevemos e que compõem a linha de base, o Maxent PL se mostrou incapaz de aprender restrições relacionadas ao fenômeno da harmonia vocálica já bem conhecidos e descritos na literatura sobre shona. Após adicionar ao aprendiz um componente (projeções) que permitia ao Maxent PL analisar apenas as sequências de vogais nas palavras que compunha a amostra de dados, obteve-se um sucesso considerável no aprendizado de restrições relacionadas à harmonia vocálica. O modelo, portanto, ofereceu suporte a teorias que propõem um nível de estrutura no qual as vogais de uma dada palavra podem ser vistas como adjacentes (independentemente de quantas consoantes estão entre elas no nível mais superficial da palavra original) (e.g. tier vocálico da fonologia autosegmental (CLEMENTS, 1976)).

Retornando, portanto, à discussão dos resultados de nossa aplicação, seguindo o princípio da linha de base indutiva, acreditamos que é importante testar se a adição de um componente de estrutura silábica ao Maxent PL leva ao aprendizado de restrições poderosas no domínio da sílaba e que já são bem descritas na literatura, como por exemplo, ocupar a segunda posição de um *onset complexo* com uma consoante diferente de *l* e *r*. Para implementar tal ideia, decidimos trabalhar especificamente com os *onsets do PB*, tal como Hayes e Wilson fizeram com o inglês. Para tanto, escrevemos um programa que quantifica o número de vezes que cada *onset do PB* aparece na amostra de dados utilizada na primeira aplicação. Podemos ver o resultado do programa na figura abaixo (o que corresponde à nossa nova amostra de dados):

Figura 04 - Amostra de dados para aplicação dos onsets

p l	32
k l	36
g l	6
f l	16
p r	304
b r	96
d r	34
t r	250
k r	84
g r	82
f r	36
p	726
t	2084
k	1214
b	378
d	1798
g	422
f	474
v	702
s	1328
z	434
x	108
jj	272
h	850
r	1002
l	832
lh	86
m	1112
n	1214
nh	106

Como vimos em 2.1.2, o Maxent PL também aceita amostras de dados que informam frequência de ocorrências. Portanto, para simular o uso de um componente silábico (neste caso, de *onsets*) no aprendizado do Maxent PL, tudo que precisamos fazer em relação à primeira aplicação foi: i) utilizar os dados acima no lugar da amostra de dados completa; ii) retirar do outro input (i.e. o inventário fonológico representado em traços distintivos) todas as vogais, pois se as vogais estivessem no inventário e não fossem computadas uma única vez na amostra de dados, o Maxent PL aprenderia fortes restrições relacionadas a elas, o que desclassificaria nossa nova aplicação como um aprendizado no escopo de *onsets do PB*, já que vogais, por definição, não ocupam tal posição na sílaba.

Aplicações testes com os *onsets* demonstraram não ser necessário determinar um número máximo de restrições a serem aprendidas ou diminuir o último intervalo da escala iterativa de acurácia (cujo padrão do software, como vimos, é $O[C_i]/E[C_i] \leq 1$). O aprendizado completo encerra ainda na primeira hora. Portanto, nossa segunda aplicação, apenas com onsets, se justifica tanto por permitir um aprendizado que encerra

naturalmente (oferecendo assim um parâmetro de comparação para nossa primeira aplicação), quanto por implementar o princípio da linha de base indutiva propostas pelos autores do Maxent PL para um programa de pesquisa em fonotática. Resumimos abaixo nossas duas aplicações:

Quadro 07 – Parâmetros das aplicações

	Aplicação 1	Aplicação 2
Amostra de dados	Palavras completas	Apenas onsets
Último intervalo de acurácia	-	$A(C_i) < 1$
Nº máximo de restrições	57	-
Operador de complementação	Sim	Sim

A gramática fonotática obtida com a segunda aplicação pode ser vista no anexo B1 (no anexo B2 disponibilizamos um arquivo análogo àquele em A2, contendo a tradução das matrizes para segmentos de cada matriz pertencente à gramática aprendida). Uma rápida leitura das restrições já revela resultados mais semelhantes aos daqueles descritos nos trabalhos em fonologia do PB, como por exemplo, as duas primeiras restrições aprendidas. Além disso, observa-se considerável gradiência de pesos em restrições que são irrelevantes do ponto de vista de uma fonologia categórica (i.e. por serem todas violadas frequentemente no léxico do PB), tais como as restrições de apenas uma matriz. Há ainda restrições inesperadas (mais especificamente $*[^\wedge + \text{dorsal}][+ \text{anterior}]$ e $[^\wedge - \text{continuant}, - \text{voice}, + \text{labial}][+ \text{anterior}]$) com pesos relativamente altos. De forma mais geral, contudo, a principal conclusão que tiramos do resultado da segunda aplicação é de que é difícil para o Maxent PL aprender restrições poderosas no escopo da sílaba sem utilizar noções de componentes estruturais postulados pela teoria fonológica.

A partir de agora nos referiremos à primeira aplicação (2.1.1) como aplicação/simulação completa e à segunda aplicação (2.1.4) como aplicação/simulação dos onsets. Por extensão, chamaremos as *gramáticas fonotáticas* resultantes dessas simulações de gramática completa e gramática dos onsets.

2.1.5 Determinando S e M

De posse das duas gramáticas, seguimos as seguintes etapas: 1) selecionamos uma amostra de 15 restrições de cada gramática; 2) para cada restrição selecionada, construímos uma não-palavra³¹ (i.e. uma *palavra* que sabemos não estar associada a significados por falantes do PB) que violasse sua respectiva restrição, e apenas ela (em sua gramática), uma única vez; 3) calculamos a *valor de Entropia Máxima* das 30 não-palavras. Abaixo descrevemos como procedemos em cada etapa.

Procedimentos em 1): Para seleção das restrições das gramáticas, realizamos uma amostragem aleatória. Enumeramos as restrições em cada gramática de acordo com a ordem que foram aprendidas (sendo 58 restrições da gramática completa e 62 da gramática dos *onsets*) e, em seguida, escrevemos um pequeno programa em Python para gerar aleatoriamente (por meio da função *randint()*) 15 números inteiros não repetidos entre 1 e 58, e 15 números inteiros não repetidos entre 1 e 62. As restrições associadas aos números gerados para cada gramática foram, então, selecionadas para compor nossa amostra.

Procedimentos em 2) e 3): nos quadros abaixo, um para cada gramática, podemos ver as *restrições* selecionadas, as não-palavras associadas a elas (as formas à esquerda na segunda coluna, e que chamamos de formas violadoras: F.V.), e os *valores de Entropia Máxima* destas não-palavras ('Maxent' no quadro). Além disso, na segunda coluna, ao lado de cada forma violadora, adicionamos ainda as formas de controle (F.C.) que serão utilizadas no experimento que descreveremos em 2.2. As formas de controle também são não-palavras, mas que, diferentemente das formas violadoras, não violam qualquer restrição de sua gramática ou violam uma restrição de peso muito baixo comparado com o peso da restrição violada por seu par³².

³¹ Este é sem dúvidas um péssimo termo para se referir ao conceito, mas preservamos a terminologia por já ser bem estabelecida na literatura. Não desejamos, portanto, qualquer interpretação de que a coisa (i.e. palavra) é palavra e não é ao mesmo tempo, deixando este tipo de lógica absurda para aqueles que aceitam a interpretação de Copenhagen na física quântica.

³² Uma exceção sendo o par *anabo-anhabo*, pois ambos também são violados por [+nasal]. Porém, justamente por ambos serem violados, resolvemos não computar tal violação no cálculo da *pontuação*.

Quadro 08 – Amostra da aplicação completa: restrição, F.V., F.C., e valor Maxent da F.V.

Restrição	F.V. – F.C.	Maxent (F.V.)
*[-cont,+voice][-appr,+ant]	bsada – bada	0.026
*[^-nasal,- high,+round][+voice][+word_boundary]	anôd – anôda	0.011
*[+nasal,-high,-back][+word_boundary]	badẽ – bade	0.012
*[-cons,+nasal][+low]	dõata - donata	0.223
*[+back][+low,+round]	tãorda – torda	0.051
*[-cons,+nasal][^nasal,-high,+round][+voice]	ãobda - auda	0.024
*[+low,-back][+son,-syll][^+high,-back]	délita – détila	0.018
*[-cont][+con,+nasal]	gnata - gonata	0.016
*[+cont,-voice,+cor][+son,-syll][^nasal,-high]	casmida - casmada	0.019
*[-high][+nasal,+high]	bãinda - baida	0.033
*[-cont,+voice][-appr,+ant]	absca - abasca	0.026
*[+cons,+son][+word_boundary]	todalh - total	0.010
*[^nasal,-high,+back][+con,+lab][+word_boundary]	bataf - batal	0.016
*[-back][+low,-back]	dié - dio	0.029
*[^+cont,-voice,+ant][+nasal][+cons,+appr]	tasonl – tasol	0.020

Quadro 09 – Amostra da aplicação dos *onsets*: restrição, F.V., F.C., e valor Maxent da F.V.

Restrição	F.V. – F.C.	Maxent (F.V.)
*[[^] -cont][-word_boundary]	pfata – fata	0.122
*[+voice,+cor]	jeda - teda	0.858
*[[^] -voice,+lab][+lat]	flado – tado	0.659
*[-voice,+dor][-word_boundary]	acstade - acastade	0.830
*[-son,-ant]	chabo – dabo	0.509
*[+nasal,+cor]	anabo - anhabo	0.858
*[+spread][-word_boundary]	hbado - baco	0.097
*[[^] -cont,-voice,+lab][+ant]	apa - óta	0.543
*[-cont][[^] -lat]	crada - rada	0.326
*[-word_boundary][-ant,+lat]	blhata – belata	0.594
*[-cont,-voice][[^] -lat]	proda - pota	0.548
*[+nasal]	maba – taba	0.675
*[+son][-word_boundary]	mdota - madota	0.088
*[+son,-ant]	araba – ataba	0.477
*[+lab][[^] +appr,+ant]	vlaco – laco	0.314

Para o cálculo do *valor de Entropia Máxima* (doravante *valores Maxent*, escrevemos um pequeno programa para facilitar a tarefa. Deixamos o programa disponível no apêndice C2 com uma breve explicação de como utilizá-lo.

Como o leitor já deve ter desconfiado, as formas violadoras dos quadros acima constituem nosso conjunto S , e os *valores Maxent* de cada um desses elementos compõem nosso M (pois como vimos no início do capítulo 1, não calculamos diretamente $P(s_1)$ e sim um valor proporcional: os *valores de Entropia Máxima*). Voltando então a denotar os *valores Maxent* de cada $s_i \in S$ como $P^*(s_i)$, temos que:

$$M = (P^*(s_1), P^*(s_2), \dots, P^*(s_{30}))$$

Entretanto, pretendemos calcular a acurácia de nossas duas *gramáticas fonotáticas* separadamente. Assim, repartiremos S em S_c (as formas violadoras do quadro da *gramática completa*) e S_o (as formas violadoras do quadro da *gramática dos onsets*), nos levando então a uma repartição também de M em M_c e M_o , cada um contendo a *pontuação* apenas de seus respectivos conjuntos, ou seja:

$$M_c = (P^*(s_1), P^*(s_2), \dots, P^*(s_{15})), \text{ para cada } s_i \in S_c$$

$$M_o = (P^*(x_1), P^*(x_2), \dots, P^*(x_{15})), \text{ para cada } x_i \in S_o$$

Quando não for necessário distinguir as *gramáticas*, continuaremos falando em S e M de maneira geral. O mesmo acontecerá com E , que, como veremos, só precisará ser repartido na hora do tratamento estatístico, uma vez que os procedimentos para obter dados fonotáticos experimentais para elementos de S_c e S_o são os mesmos. Nosso trabalho, portanto, possui o mesmo problema de pesquisa, com a única diferença de que iremos resolvê-lo para duas *gramáticas fonotáticas* distintas aprendidas pelo Maxent PL.

2.2. Construindo E

Chegamos ao ponto do trabalho no qual extrapolaremos o escopo da linguística teórica, e seus métodos matemáticos, e lidaremos experimentalmente com o fenômeno linguístico que motiva o tipo de modelagem visto até então. Sendo o Maxent PL uma teoria científica da mente humana³³, os postulados assumidos até então são postulados sobre a mente. Assim, assume-se a existência cognitiva da *fonotática* e das *palavras* (além de seus componentes) com suas respectivas *boa-formações*³⁴. Com isso, nosso objetivo pode ser visto como a tentativa de eliciar, já que não temos como observar diretamente fenômenos cognitivos, a *boa-formação* que falantes-ouvintes da língua de interesse (PB, em nosso caso) associa às *palavras* de um dado conjunto (e.g. o conjunto *S*). Note-se, portanto, que, obviamente, mas não trivialmente, experimentos em linguística (assim como em toda psicologia) partem do postulado de que a observação de certos dados eliciados dos sujeitos podem servir como evidência psicológica, ou, fazendo um uso mais específico desse postulado para nosso trabalho, certas observações eliciadas podem evidenciar a *boa-formação* de formas linguísticas³⁵.

O método padrão adotado em linguística formal para se tentar evidenciar a *boa-formação* da mente dos falantes-ouvintes de uma dada língua consiste dos chamados julgamentos de aceitabilidade. Esses julgamentos consistem basicamente, no caso de experimentos fonotáticos, da resposta dada por estes falantes-ouvintes a perguntas como “quão aceitável seria essa *palavra* para você como parte da sua língua?” ou “quão natural essa *palavra* soa para você?”³⁶ quando apresentados a estímulos (i.e. via áudio ou escrita) que representam as *palavras*. Portanto, fazendo uma relação com o parágrafo anterior, utilizam-se os seguintes termos no contexto experimental (BARD et. al., 1996, p.33): i) *gramaticalidade*: a *boa-formação* da forma no nível cognitivo em si (a qual assumimos existir e não temos acesso direto); ii) *aceitabilidade*: como a forma (representada, claro,

³³ Mais precisamente, de certos tipos de objetos e processos mentais (aqueles de natureza fonológica).

³⁴ Assim como em outras áreas do conhecimento (e.g. a matemática e a lógica), postulados como estes, que compõem os fundamentos da linguística formal, tem como base intuições simples da mente humana. Seja em uma perspectiva categórica ou probabilística de “fonotática”, nos parece inescapável a ideia de que utilizamos sequências de objetos abstratos para compor as “palavras” de uma língua natural e que nem todas as permutações de tais objetos parecem pertencer a essa língua, ou pelo menos, não parecem “pertencer” no mesmo nível.

³⁵ Afinal, são essas observações e a própria intuição do pesquisador enquanto falante nativo de uma língua que motivam a postulação da existência dos objetos cognitivos que mencionamos neste parágrafo.

³⁶ No caso de um estímulo sonoro que representa a *palavra*.

por um estímulo) é percebida pelo falante-ouvinte em relação à pergunta feita; iii) *julgamento de aceitabilidade*: a resposta dada pelo falante ouvinte ao questionamento do linguista com base em sua aceitabilidade da *palavra* apresentada.

Os *julgamentos de aceitabilidade*, portanto, equivalem ao que chamamos de dados eliciados dos sujeitos que podem servir como evidência psicológica. Eles constituem a única parte observável para o linguista. Entretanto, embora resida nesses julgamentos a esperança de se obter resultados sobre a *gramaticalidade*, é quase evidente, a partir das três definições acima, que se reconhece as limitações do método diante da não equivalência entre *aceitabilidade* e *gramaticalidade*, tendo em vista os diversos fatores que podem influenciar a *aceitabilidade* além da *gramaticalidade* em si. Tais limitações são, de fato, reconhecida pelos linguistas, como vemos abaixo na seguinte citação de Bard et. al. (1996, p.33):

Of course, acceptability judgments, like other manifestations of linguistic performance, need not be one-to-one reflections of grammaticality. First, it is always possible that the subject is not reporting directly on grammaticality but is responding to any number of other features of the stimulus (Botha 1973, Quirk & Greenbaum 1970). Impressions of acceptability may be based, for example, on estimated frequency of usage, on conformity to a prescriptive to a prescriptive norm or a prestigious register, or on degree of semantic or pragmatic plausibility.

Naturalmente, quando as *palavras* apresentadas são formas que não pertencem ao léxico do PB e que se sabe não haver significados associados a elas (i.e. o que chamamos na subseção anterior de não-palavras), os exemplos dados na citação de possíveis fatores que influenciam a *aceitabilidade* são extintos. Daí nossa escolha, como na grande maioria de outros experimentos fonotáticos, por não-palavras. Todavia, apesar de ser uma escolha superior, isto não significa que não existam outros fatores que influenciem a aceitabilidade de não-palavras; para mencionar alguns candidatos, temos a semelhança entre a não-palavra com uma palavra do léxico do PB, ou da interpretação da pergunta em si.

2.1 Estimação de magnitude

Outra questão de extrema importância na utilização do método diz respeito à escala utilizada para medir o *julgamento de aceitabilidade*. Pela nossa definição de *fonotática* adotada desde o final da introdução, a *boa-formação* de uma *palavra* pode assumir um dentre infinitos valores (qualquer real entre 0 e 1). Portanto, uma escala de infinitos valores deve estar à disposição do falante-ouvinte para o *julgamento de aceitabilidade* se pretendemos de fato utilizar tais julgamentos como maneira de evidenciar a *boa-formação*.

Isto significa que escalas ordinais comuns, como por exemplo, escalas de 0 a 10 (ou com qualquer número natural n de opções) não permite toda a expressão de gradiência necessária. É fácil investigar as limitações dessas escalas. Suponha que em um dado experimento, um sujeito atribuiu o valor 5 de *aceitabilidade* (em uma escala de 0 a 10) a um primeiro estímulo e depois 10 a um segundo estímulo (por considera-lo duas vezes mais bem-formado que o primeiro). Em seguida, suponha que foi apresentado um terceiro estímulo que o sujeito considerou ainda mais bem-formado que o segundo, então qual valor ele atribuirá a esse terceiro estímulo? A resposta será provavelmente 10 também, mas fica evidente que em um caso como esses, o valor do *julgamento de aceitabilidade* foi restringido por uma limitação da escala e não da *aceitabilidade* (e da gradiência fonotática em si), e essa informação seria perdida, uma vez que o linguista só tem acesso ao *julgamento de aceitabilidade* e não à *aceitabilidade*.

Essas simples e essenciais considerações sobre as escalas utilizadas em *julgamentos de aceitabilidade* formam a principal motivação para que pesquisadores introduzissem à linguística na década de 90 (BARD et. al., 1996) um tipo de experimento advindo da psicofísica (STEVENS, 1957), a estimação de magnitude, e o qual utilizaremos nesta dissertação. A respeito deste tipo de experimento, Bard et. al. (1996, p.40) explicam:

Magnitude estimation was developed to provide better than ordinal scales for measuring impressions of physical continua (Stevens, 1956). As originally applied to the direct estimation of brightness or loudness, magnitude estimation in its simplest version requires the subject to associate a numerical judgment with a physical stimulus (see Stevens 1975 for a review). Once the initial stimulus, or modulus, is presented and a number associated with it by the experimenter or subject, the subject assigns to each successive stimulus a number reflecting the relationship between that stimulus and the modulus.

Essa relação entre os estímulos subsequentes ao módulo e o módulo devem então ser estimada da seguinte maneira (BARD et. al., 1996, p.40):

Subjects are explicitly instructed to reflect perceived ratios in their judgments: a stimulus that appears to be 10 times as bright as the first is to be given a number 10 times the original number. One that seems one-third as bright is given a number one third the size. However bizarre they may find the task at first, normal adults can reliably perform it for a large number of physical continua.

A propriedade de maior importância para nosso trabalho e que devemos notar em relação à estimação de magnitude é que a escala possui gradiência ilimitada. Para ilustrar com um caso extremo, se o sujeito julga um estímulo 10^{10} vezes maior ou 10^{-10} menor que o módulo em relação à propriedade que está sendo observada (brilho, volume, tamanho, ou, já generalizando o experimento para domínios mais abstratos do que propriedades físicas: *boa-formação*), basta atribuir, respectivamente, um número real 10^{10} vezes maior ou 10^{-10} menor que o número atribuído inicialmente ao módulo. Uma segunda propriedade importante diz respeito aos resultados já obtidos na psicofísica, a qual adiciona maior credibilidade a esse tipo de experimento. De acordo com Bard et. al. (1996, p.41):

Although there is dispute about the generality of the finding (see Poulton 1986, 1989 for a critique), when the subject's estimates of magnitude (or group geometric mean estimates or medians) are plotted in log-log coordinates against the physical dimension, the points tend to follow a straight line with a slope characteristic of the physical property being assessed. The straight line in log-log coordinates means that equal ratios on the physical dimension give rise to equal ratios of judgments. In judgments of brightness, for example, every time the stimulus energy doubles, the subjective brightness becomes 1.5 times larger. In judgments of line length, on the other hand, the function is steeper: doubling physical line length doubles subjective line length as well. The characteristic relationship is reflected in the value of this slope, called *b* or *B*.

Se chamarmos de *R* a variável dos valores atribuídos aos estímulos (compondo o eixo *y*) e de *S* a variável dos valores dos estímulos medidos em relação a sua propriedade física (compondo o eixo *x*), o que os resultados acima nos diz matematicamente é que o gráfico formado pelos logs dos valores no eixo *y* e pelos logs dos valores no eixo *x* formam uma reta, ou seja:

$$\log R = c + b \log S$$

Onde c e b são constantes nos reais.

Assim, segue que

$$e^{\log R} = e^{c+b \log S}$$

$$e^{\log R} = e^c \cdot e^{(\log S)^b}$$

$$R = e^c \cdot S^b$$

Se chamarmos a constante e^c de k , chegamos às formas apresentadas por Bard et. al. (1996, p.41) para tais equações:

$$\log R = \log k + b \log S$$

$$R = k \cdot S^b$$

Assume-se ainda que $\psi = R$, onde ψ representa a magnitude subjetiva do estímulo (i.e. a magnitude tal como percebida cognitivamente pelo indivíduo), o que significa que, assim como na linguística, é postulado na psicofísica que a resposta dada pelo sujeito (nesse caso, o número atribuído ao estímulo) equivale à percepção da magnitude que se deu no nível psicológico (analogamente à equivalência que esperamos haver entre *aceitabilidade* e o *julgamento de aceitabilidade* em si). Estas equações que representam relações psicofísicas, de fato, são conhecidas como as leis exponenciais (STEVENSON, 1957 apud BARD et. al., 1996, p.41). Tais resultados, cujos padrões levam a postulação de leis sobre a relação entre certas propriedades físicas e percepções cognitivas das mesmas, foram encontrados utilizando a estimação de magnitude como método.

Diante do exposto até então, portanto, acreditamos haver justificado e introduzido claramente a escolha de nosso tipo de experimento para realização dos *julgamentos de aceitabilidade* fonotáticos. Destarte, a proposta é generalizar o método da estimação de magnitude para além do domínio da psicofísica, ou seja, para estimar uma propriedade que não possui um correlato físico conhecido (e.g. tamanho de retas, intensidade de brilho), e sim uma propriedade abstrata do estímulo: a *boa-formação*. De maneira geral e análoga ao que descrevemos mais acima, o procedimento consiste em apresentar um primeiro estímulo (representando uma *palavra*) que servirá como módulo, atribuir um

valor a este módulo para representar sua *aceitabilidade*, e pedir que os sujeitos atribuam um *julgamento de aceitabilidade* aos estímulos subsequentes com base no valor atribuído ao módulo. Portanto, se o sujeito atribuiu um valor m qualquer ao módulo em relação à sua *aceitabilidade* e considera um estímulo subsequente três vezes “mais aceitável/natural” como possível forma de sua língua, deve-se, então, atribuir o valor $3m$ ao estímulo em questão. Um dos autores do Maxent PL já aplicou a estimação de magnitude a testes fonotáticos de *aceitabilidade*, e nosso experimento se baseia neste trabalho (HAYES&WHITE, 2012).

2.2 Estimação de magnitude aplicada à fonotática

2.2.1 Participantes

Participaram do experimento 12 estudantes de graduação e pós-graduação da Universidade Federal da Paraíba. A população inteira de nosso experimento consiste de todos os falantes nativos do PB, e desconhecemos qualquer variável relacionada aos indivíduos que influencie significativamente nossa variável dependente neste tipo de experimento. Assim, realizamos uma amostra por conveniência, selecionando os 20 primeiros estudantes a se voluntariarem.

2.2.2 Materiais

Para aplicação da magnitude de estimação, são necessários um mínimo de dois materiais: i) o conjunto de estímulos a serem apresentados; ii) um registro dos números atribuídos aos estímulos. Nosso conjunto de estímulos já foi apresentado nos quadros 08 e 09, pois consiste, naturalmente, das *palavras* de nosso conjunto S e de formas de controle para cada *palavra* em S (todos definidos e exibidos na segunda coluna dos quadros mencionados, em 2.1.5).

Para apresentar os estímulos aos participantes, registrar os valores da estimação e computar nossa estatística descritiva, escrevemos um programa que apresentaremos nos procedimentos. Além disso, o programa oferece um componente de áudio para complementar as representações gráficas dos quadros acima enquanto estímulo para os participantes. Assim como Hayes e White (2012), realizamos ainda uma etapa de calibragem com os participantes antes do experimento fonotático. Nesta etapa, realizamos

a estimação de magnitude com tamanho de retas para assegurar que a tarefa de estimação a ser realizada fosse bem compreendida e que os participantes estivessem de fato prestando atenção. Descrevemos os procedimentos na próxima subseção.

2.2.3 Procedimentos

Apresentaremos primeiramente a etapa de calibragem, e, em seguida, nosso experimento de estimação de magnitude para julgamento de aceitabilidade fonotático.

2.2.3.1 Etapa de calibragem

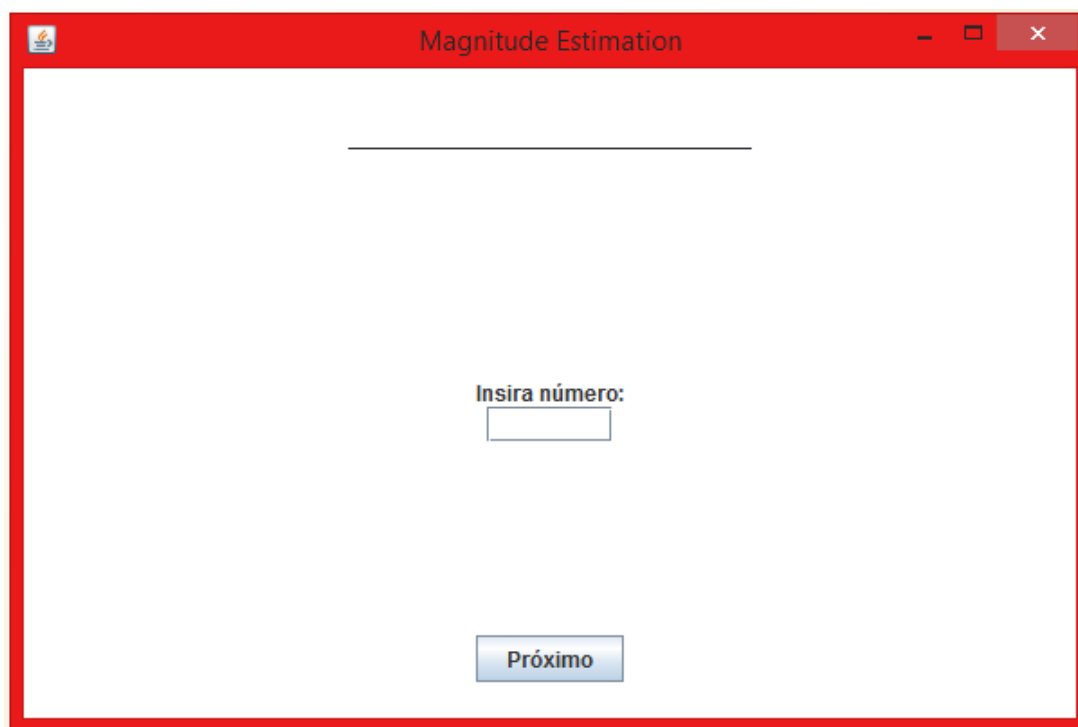
De acordo com Bard et al. (1996, p.41), as pesquisas em psicofísica mostram que:

In judgments of brightness, for example, every time the stimulus energy doubles, the subjective brightness becomes 1.5 times larger. In judgments of line length, on the other hand, the function is steeper: doubling physical line length doubles subjective line length as well.

Tal relação é encontrada, como vimos, em coordenadas log-log e significa que $b = 1$ nas leis exponenciais que exploramos na subseção anterior. Esta forte e direta correlação significa que experimentos de estimação de magnitude com tamanhos de retas são provavelmente um bom recurso para verificar se participantes de um experimento compreendem a tarefa em mãos, além de familiarizá-los com os procedimentos antes de aplicar um segundo experimento desse tipo. Isto posto, realizamos um experimento de estimação de magnitude de retas antes de aplicar nosso experimento fonotático.

Nossa reta módulo foi construída com 5cm, e selecionamos seis outras retas cujas razões em centímetros em relação ao módulo eram 0.5, 2, 1.5, 0.3, 0.2, e 2.5. Em nosso programa, iniciamos então o experimento apresentando a seguinte tela contendo o módulo e uma caixa para atribuição de valores reais:

Figura 05 – Programa de estimação de magnitude de retas (tela inicial da interface do usuário)

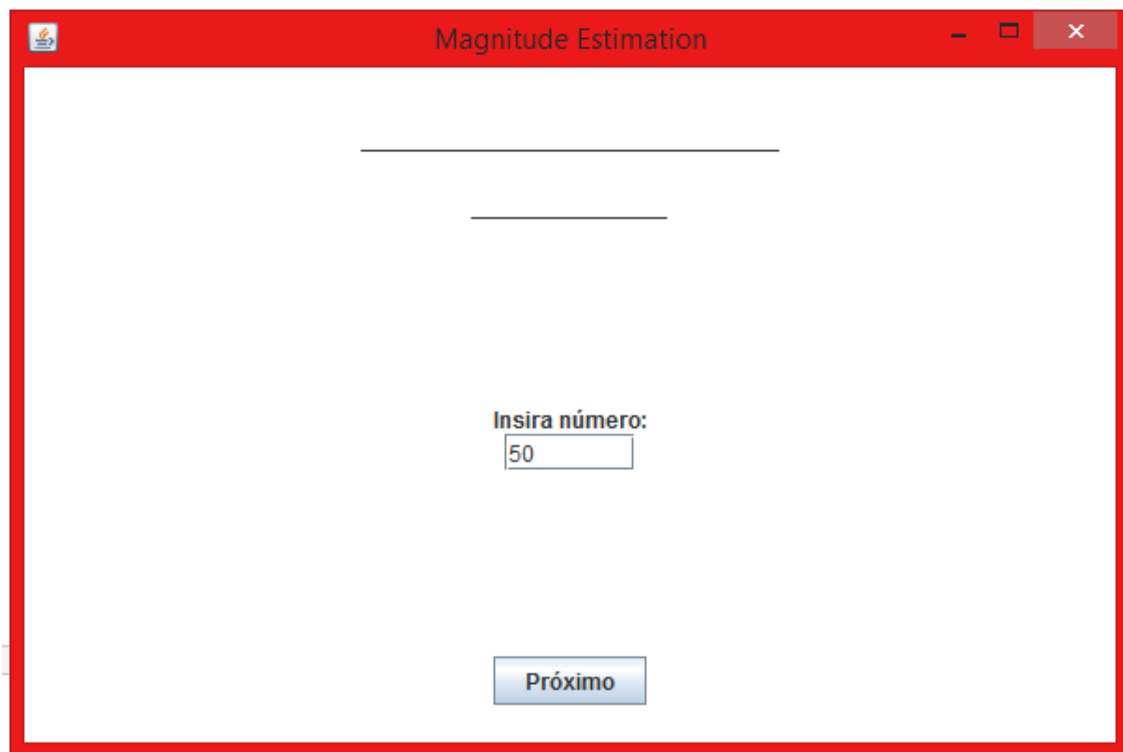


Em seguida, explicamos ao participante que o valor 100 será atribuído a esta reta apenas por ser um número fácil de multiplicar e dividir mentalmente, mas que poderíamos haver escolhido qualquer outro valor (o que é verdade). Na sequência, explicamos também que a partir de então o número 100 representará o tamanho da reta e, com base nisso, o tamanho das retas subsequentes deverá ser estimado em comparação com a reta de referência. Isto deve ser feito de maneira que, se o participante perceber a reta subsequente como 3 vezes maior que o módulo, ele deverá atribuir 300 à reta, e assim por diante. A reta módulo será exibida na tela durante todo o experimento, enquanto as retas estimadas mudam quando o participante pressionar o botão <Próximo>. Explicamos ainda que não é possível voltar a retas anteriores uma vez que o botão foi pressionado, mas se desejado podemos reaplicar o experimento a partir da parte necessária³⁷.

³⁷ Isto é possível porque, como veremos em mais detalhes mais adiante, o programa salva todos os resultados obtidos para cada reta estimada, então podemos fechar, reabrir o programa e deixar na reta desejada pelo participante sem perder informação.

Abaixo mostramos a tela obtida após pressionar o botão <Próximo> pela primeira vez para ilustrar nossa descrição:

Figura 06 – Programa de estimação de magnitude de retas (interface do usuário para estimação)



Como podemos ver, uma nova reta aparece abaixo do módulo, e na caixa colocamos a estimativa que corresponderia exatamente à razão da medição física (pois a reta debaixo mede 2.5 cm), mas na verdade, assim que o botão <Próximo> foi pressionado a caixa havia sido esvaziada. Ao pressionar próximo, o valor escolhido já é escrito no terminal de nosso programa juntamente com o valor da estimativa dividido por 100, ou seja, a razão entre a estimativa e o módulo. Desta maneira, ao final do experimento, obtemos um output como no exemplo abaixo (resultado do participante 01):

Figura 07 – Output da estimação de retas do participante 01

```

100
50
200
150
40
10
250
(1, 1.0)
(0.5, 0.5)
(2, 2.0)
(1.5, 1.5)
(0.3, 0.4)
(0.2, 0.1)
(2.5, 2.5)

```

As sete primeiras linhas mostram os valores atribuídos à cada linha na estimação (incluindo o módulo) na ordem em que foram apresentadas no programa, e as sete últimas mostram pares ordenados nos quais os números da esquerda mostram a razão entre o tamanho físico das linhas e do módulo em centímetros, enquanto que os números da direita são a razão obtida na estimação. No exemplo acima, podemos ver uma forte correlação, que não chega a 1 quando medida em Pearson, mas que ultrapassa 0.9. Para realizar nosso experimento fonotático, todos os participantes tiveram que exibir resultados semelhantes nessa etapa de calibragem como demonstração de que compreenderam o método do experimento e de que são capazes de estimar ao menos propriedades simples. Passamos agora ao experimento fonotático.

2.2.3.2 Determinando *E*

Chegamos, enfim, na seção em que descreveremos os procedimentos utilizados para construir nosso conjunto *E*. Passada a calibragem, os participantes são informados de que irão realizar os mesmos procedimentos, mas agora avaliando uma propriedade mais abstrata do que o tamanho de retas. Apresentamos então a tela inicial do novo experimento e a partir dela continuamos as instruções:

Figura 08 – Programa de estimação de magnitude fonotática (tela inicial da interface do usuário)



Explicamos que agora estaremos avaliando palavras inventadas (não-palavras) ao invés de retas, e a palavra <blada> servirá como nossa referência dessa vez e todas as outras palavras serão comparadas com ela. Informamos, então, que a questão agora é “Quão aceitável essa palavra soaria para você como palavra do PB? Parece algo que soaria mais estranho como uma palavra da sua língua ou algo que pertenceria bem a ela?”. Novamente atribuímos 100 ao módulo (pelos mesmos motivos da etapa de calibragem) para representar quão “boa” esta palavra soaria como palavra do PB, e pedimos que os participantes atribuam valores às palavras subsequentes tendo em mente sempre o seguinte: “se ‘blada’ recebeu 100 em relação à pergunta feita, quanto esta outra palavra merece?” Em caso de dúvidas criamos algum exemplo extremo na hora (e que não está

no conjunto de estímulos do experimento) para checar se o participante atribuiria um valor menor.

Após explicar a tarefa, informamos que, de maneira análoga ao experimento das retas, <blada> permanecerá sempre escrita na tela e que ao pressionar <próximo> as palavras seguintes serão apresentadas. Além da representação grafêmica, toda palavra vem acompanhada por uma pronúncia da mesma, a qual pode ser ouvida ao pressionar o botão com o símbolo de “tocar”, tal como indicado na tela acima. Sugerimos que a opção auditiva seja utilizada quando a forma escrita de uma dada palavra causar dúvidas sobre como ela seria pronunciada ou para confirmar a própria pronúncia imaginada/realizada pelo participante. O áudio pode ser tocado quantas vezes o participante desejar, bastando pressionar o botão para tocá-lo.

Assim como para as retas, não é possível retornar a uma palavra depois de pressionar <Próximo>. Porém, o programa é robusto, e só passa para a palavra seguinte se houver um número na caixa de input; assim, o participante só precisa se preocupar em fazer a estimativa que realmente deseja antes de pressionar o botão, mas não corre o risco de clicar mais de uma vez e pular alguma palavra. Ademais, assim como para as retas, mesmo que tenhamos que fechar o programa e rodá-lo novamente, é possível deixar o participante refazer o experimento a partir de qualquer ponto sem que percamos resultados do que havia sido feito antes.

Encerramos, assim, as instruções para os participantes. Na tela abaixo, visualizamos como são as telas subsequentes à primeira que contém apenas o módulo:

Figura 09 – Programa de estimação de magnitude fonotática (interface do usuário para estimação)



A partir da segunda tela (após digitar 100 para ‘blada’ e pressionar o botão <Próximo>), o participante não tem mais acesso à pronúncia de blada, apenas à sua representação grafêmica, pois o botão ‘tocar’ agora reproduz a pronúncia do estímulo atual. Concluímos, dessa maneira, os procedimentos utilizados para aplicar o experimento com os participantes a partir de nosso programa. Agora esclareceremos algumas questões técnicas importantes, antes de finalizar exibindo o output de um dos participantes e como construímos *E* a partir desses resultados.

As gravações que compõem o áudio foram realizadas por uma falante nativa do PB com formação em fonética e fonologia. Para gravar, utilizamos o software livre Audacity e um headset Logitech H390. Ao final, utilizamos as configurações padrões do software para equalizar o volume e reduzir o ruído. Como este não é um trabalho em fonética acústica e nosso interesse é apenas oferecer um complemento auditivo inteligível aos estímulos de nosso experimento, consideramos os procedimentos de gravação satisfatórios, assim como o resultado final.

Os estímulos foram apresentados na mesma ordem para todos os participantes. Para estabelecer a ordem em que os estímulos seriam apresentados, utilizamos um método semelhante àquele empregado em 2.1.5 para seleção da amostra de *restrições* das gramáticas fonotáticas; desta vez, enumeramos as 60 palavras (as 30 de *S* e seus pares de

controle) de 1 a 60 e escrevemos um pequeno programa para gerar uma sequência de 60 números sem repetições entre 1 e 60 (i.e. basicamente embaralhar os números). Entretanto, adicionamos uma restrição extra na ordem: uma forma violadora e seu par de controle deveriam ter uma distância maior ou igual a 3, ou seja, se uma forma como *absca* foi selecionada como a 10ª palavra da sequência de estímulos, *abasca* não poderia ocupar as posições de número 8, 9, 11 ou 12.

Assim como para as retas, o output de nosso programa consiste da razão entre o valor de estimação atribuído pelo participante e o valor do módulo para cada estímulo. Além disso, o programa escreve também o log da razão e o nome da palavra associada a ambos os valores. Para ilustrar, mostramos abaixo parte do output obtido no experimento com o informante 03:

Figura 10 – Output da estimação fonotática do participante 03

```
mdota ratio      0.1
mdota log      -2.3025850929940455
anôda ratio     2.0
anôda log      0.6931471805599453
bataf ratio     0.05
bataf log      -2.995732273553991
hbado ratio     0.05
hbado log      -2.995732273553991
gnata ratio     0.02
gnata log      -3.912023005428146
bade ratio      1.0
bade log       0.0
```

Temos três opções úteis para construir E_c , ou seja, três opções que nos permite calcular médias geométricas. Definamos a seguinte notação para todo $s \in S$ (o conjunto das 30 palavras em nossa amostra):

- i. $estimação_i(s)$ é o valor que o participante i atribuiu a s , com $1 \leq i \leq 12$;
- ii. $razão_i(s)$ é a razão entre $estimação_i(s)$ e o valor do módulo (100 em nosso caso);
- iii. $\log(estimação_i(s))$ é o log de $estimação_i(s)$;

Como os logs das razões são em grande parte valores negativos, teríamos que excluir ou manipular grande parte dos dados para poder calcular médias. Por outro lado, para a grande maioria dos casos, $estimação_i(s)$ (para qualquer s ou i) foi menor que a base do logaritmo natural (ou seja, os valores foram raramente negativos), significando que podemos utilizar $\log(estimação_i(s))$ se ignorarmos alguns poucos dados. Definimos então a seguinte notação para apresentar na sequência as possibilidades para construção de E

- iv. $\mu[estimação(s)]$ é a média geométrica de todos os valores $estimação_i(s)$;
- v. $\mu[razão(s)]$ é a média geométrica de todos os valores $razão_i(s)$;
- vi. $\mu[\log(esitmação(s))]$ é a média geométrica de todos os valores $\log(estimação_i(s))$.

Podemos considerar então as seguintes definições para E :

Para todo s_j , com $1 \leq j \leq 30$ (enumeração dos elementos de S), temos:

$$E_{estimação} = (\mu[estimação(s_1)], \dots, \mu[estimação(s_{30})]);$$

$$E_{razão} = (\mu[razão(s_1)], \dots, \mu[razão(s_{30})]);$$

$$E_{log} = (\mu[\log(esitmação(s_1))], \dots, \mu[\log(esitmação(s_{30}))]).$$

Como os valores de $estimação_i(s)$ são obtidos simplesmente escalonando $razão_i(s)$ por 100, do ponto de vista da correlação ambos $E_{estimação}$ e $E_{razão}$ são equivalentes, pois a correlação Pearson não é sensível à escala. Preferimos utilizar $E_{razão}$ para apresentar outras estatísticas (além da correlação) em uma escala mais compacta. Por outro lado, a transformação logarítmica utilizada para construir E_{log} levará a resultados diferentes. Portanto, calcularemos a correlação para ambos $E_{razão}$ e E_{log} , para que possamos comparar o efeito das transformações. Por fim, assim como fizemos com M , $E_{razão}$ e E_{log} serão repartidos para calcular a correlação de nossas duas aplicações

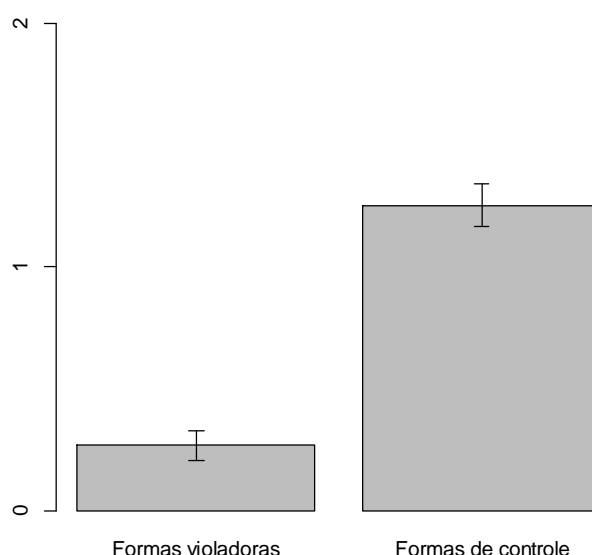
separadamente, ou seja, $E_{razão}$ será particionado em $E_{C\ razão}$ (o subconjunto de $E_{razão}$ contendo as formas da aplicação completa) e $E_{O\ razão}$ (o subconjunto de $E_{razão}$ contendo as formas da aplicação dos onsets), e, similarmente, teremos $E_{C\ log}$ e $E_{O\ log}$. Isso significa que calcularemos a correlação em 4 casos distintos.

No capítulo seguinte, apresentaremos os resultados obtidos para ambas as aplicações e finalizaremos com alguns comentários. Antes de encerrarmos este capítulo, contudo, gostaríamos de expressar que técnicas estatísticas como a utilização de escalas logarítmicas e da utilização de médias geométricas (em detrimento de aritméticas) não são idealizações e manipulações triviais dos dados (muito menos as únicas opções), e, apesar de serem raramente motivadas ou discutidas em trabalhos científicos que as utilizam (por motivos variados, tais como foco, falta de espaço, por já serem técnicas bem estabelecidas e compreendidas por uma dada comunidade de cientistas, ou até mesmo por vaidade ou falta de honestidade intelectual). Infelizmente, faremos parte deste grupo, mas alertamos os leitores sobre a importância dos fundamentos lógico-filosóficos da teoria da probabilidade, que se encontra na base desse tipo de manipulação matemática que permeia toda a ciência empírica. O mesmo se aplica aos conceitos estatísticos que apresentaremos no capítulo 3 e à própria escolha da correlação Pearson. Como referência dos fundamentos da lógica probabilística, indicamos a obra de E.T. Jaynes (2003).

CAPÍTULO III – Resultados

No capítulo II, descrevemos a metodologia empregada na obtenção dos dados necessários para resolver nosso problema de pesquisa, e, agora, apresentaremos os resultados, alguns comentários sobre os mesmos, e concluiremos o texto. Todos os cálculos estatísticos e gráficos deste capítulo foram feitos utilizando o ambiente de desenvolvimento integrado gratuito R (R CORE TEAM, 2015). Antes de apresentar as correlações, começaremos comparando as médias das formas violadoras com as formas de controle de cada aplicação. Considerando a aplicação completa, obtivemos o seguinte:

Gráfico 02 - Aplicação completa: médias das razões F.V. - F.C.



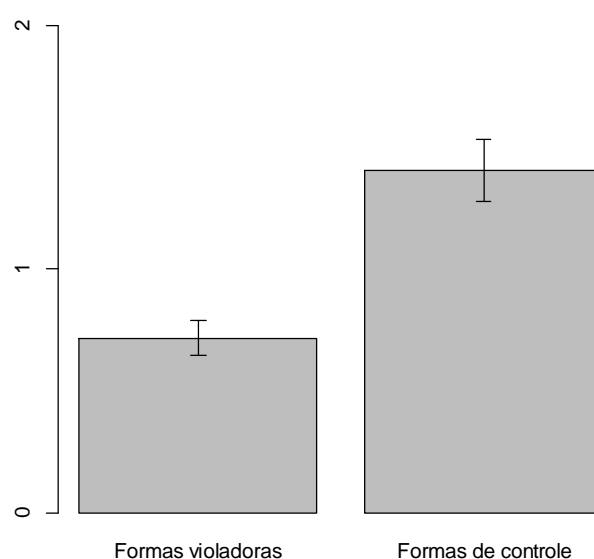
Como pode ser observado no gráfico, no geral as formas de controle receberam estimações consideravelmente mais altas ($\mu = 0.26$) que seus pares violadores ($\mu = 1.25$). O gráfico exibe ainda o erro padrão das médias. Para as formas violadoras, encontramos $SE = 0.05$ (com $SD = 0.80$)³⁸ e para as formas de controle $SE = 0.08$ (com $SD = 1.11$). Assim, se os participantes realizaram a estimaco de maneira satisfatria (i.e. seus *jugamentos de aceitabilidade* refletiram bem a *boa-formaco* das formas), os dados

³⁸ Denotaremos o erro padro por SE (do ingls *Standard Error*) e o desvio padro por SD (do ingls *Standard Deviation*).

indicam que o modelo acerta em atribuir pesos baixos a restrições que violam as formas de controle ou nem aprender tais restrições. Conversamente, se o modelo possui uma alta acurácia em relação à realidade linguística, parece que os participantes compreenderam bem a tarefa da estimação de magnitude, como a etapa de calibragem pareceu indicar.

Os resultados com a aplicação dos onsets foram parecidos, exibindo novamente uma média mais baixa às formas violadoras:

Gráfico 03 - Aplicação dos onsets - médias das razões: F.V. - F.C.



Notamos, contudo, um aumento considerável na média das formas violadoras ($\mu = 0.71$, $SE = 0.07$, $SD = 0.96$), e acreditamos que isso se deva à existência de restrições como $*[+nasal,+cor]$ e $*[^{-cont,-voice,+lab}][+ant]$ que são violadas por palavras como *anabo* e *apa*, mas não por *anhabo* e *óta*. Nossa intuição e o fato de *n* e *p* serem frequentemente violados no corpus que utilizamos nos fez estranhar o aprendizado destas restrições em conjunção com o não aprendizado de uma restrição especificamente para *j* e *t*. Ainda assim, as formas de controle exibiram uma média mais alta ($\mu = 1.40$, $SE = 1.12$, $SD = 1.73$). Na tabela abaixo, resumimos também os resultados obtidos quando utilizamos a média dos logs das estimações (retirando os logs negativos dos dados) ao invés da média das razões:

Tabela 02 - Aplicação completa – Médias dos logs das estimações: F.V. - F.C

	<i>Formas violadoras</i>	<i>Formas de controle</i>
<i>M</i>	3.55	4.77
<i>SE</i>	0.08	0.05
<i>SD</i>	1.10	0.70

Tabela 03 - Aplicação dos onsets – Médias dos logs das estimações: F.V. - F.C

	<i>Formas violadoras</i>	<i>Formas de controle</i>
<i>M</i>	4.25	4.90
<i>SE</i>	0.07	0.05
<i>SD</i>	0.96	0.75

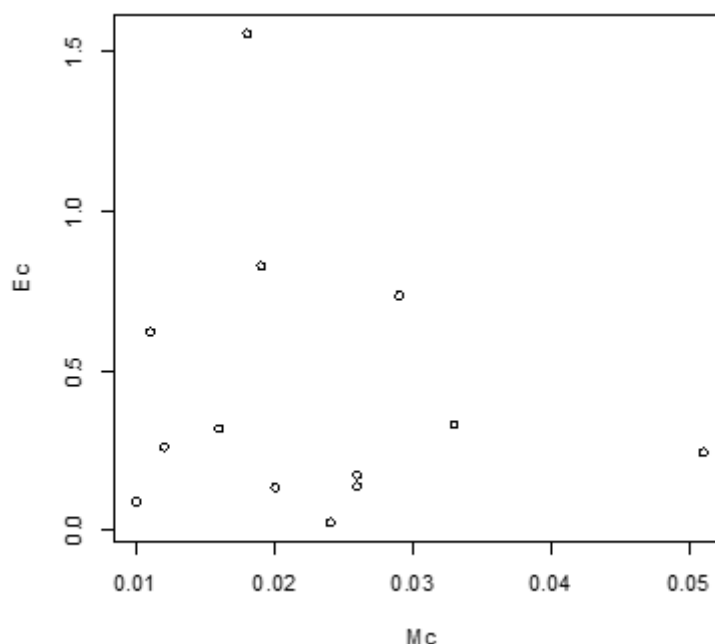
Inspeção mostra que os resultados são semelhantes e corroboram com a performance satisfatória dos participantes. Assim, passamos, finalmente, às correlações. No quadro abaixo apresentamos os resultados obtidos para a aplicação completa:

Tabela 04 - Correlações da aplicação completa

$cor(M_C, E_{C\text{ razão}})$	-0.11
$cor(M_C, E_{C\text{ log}})$	-0.02

Substituindo ainda os *valores de Entropia Máxima* que compõem M_C pelos valores de *pontuação* de suas formas, obtém-se -0.12 e -0.03 em detrimento de -0.11 e -0.02, respectivamente. Logo, independente das transformações utilizadas para E e M , o valor da correlação é muito próximo a zero, o que significa a não existência de uma dependência linear entre as previsões do modelo e os dados experimentais. Mostramos abaixo a distribuição dos dados em um gráfico de $E_{C\text{ razão}}$ em relação a M_C para ilustrar geometricamente a baixa correlação linear (removemos apenas um outlier com valor 0.22 em M_C para melhorar a visualização. Na correlação a remoção deste outlier teve um efeito insignificante, por isto nem o consideramos na discussão).

Gráfico 04 – Distribuição de $E_{C\text{razão}}$ em relação a M_C



Postulamos dois principais motivos para a ineficácia do modelo em relação à essa aplicação. O primeiro é que ao fixar um número máximo de restrições para o Maxent PL aprender, ao invés de tomar a escala de acurácia como parâmetro único para encerramento do aprendizado, podemos acabar com pesos significativamente diferentes do que o Maxent PL aprenderia. O leitor pode observar isto fazendo testes com o software, determinando números cada vez maiores de restrições (e.g. 10, depois 15, depois 20) a serem aprendidas e verificando que apesar das primeiras restrições serem quase sempre as mesmas, seus pesos variam com o tamanho final da gramática. Portanto, de fato nunca soubemos (até mesmo para o último intervalo de acurácia fixado em 0.3) quantas restrições seriam de fato aprendidas pelo Maxent PL em nossa aplicação completa e qual o efeito final disso nos pesos obtidos.

Outro possível motivo, que provavelmente atua em conjunção com o primeiro, é o extrapolamento da linha de base (i.e. as limitações do modelo simples de GU que foi adotado). Seguindo a proposta de Hayes e Wilson (2008, p.381) de linha de base indutiva para um programa de pesquisa em fonotática, tentamos aprender restrições para toda fonotática segmental do PB apenas com o aparato da Fonologia Linear (CHOMSKY&HALLE, 1968) e sem componentes da estrutura silábica. Como discutimos em 2.1.4., Os resultados da aplicação completa demonstraram ineficácia no aprendizado de restrições relacionadas a onsets, codas, e até mesmo que se mantivessem no domínio da sílaba do PB. O fato da aplicação dos onsets demonstrar resultados mais

satisfatórios nesse sentido reforça a hipótese de que na aplicação completa o aprendiz não possuía os componentes necessários para aprender os padrões fonotáticos mais gerais e significativos da língua, sendo limitado a restrições menos expressivas que encontram pouco respaldo experimentalmente.

É interessante considerar que se tivéssemos obtido uma forte dependência linear entre as previsões da aplicação completa e os dados experimentais, isso abriria caminho para uma possibilidade intrigante: que existe um número considerável de padrões fonotáticos significativos no nível segmental que extrapolam o domínio da sílaba, indo de encontro ao que estamos acostumados a pensar e fazer em fonologia. Entretanto, nosso resultado parece ressaltar a importância da estrutura silábica na boa-formação das palavras e de nossa intuição acerca da estranheza de várias restrições aprendidas na aplicação completa. Apesar disso, consideramos importante que estudos em larga escala sejam realizados para podermos fazer asserções probabilísticas mais embasadas.

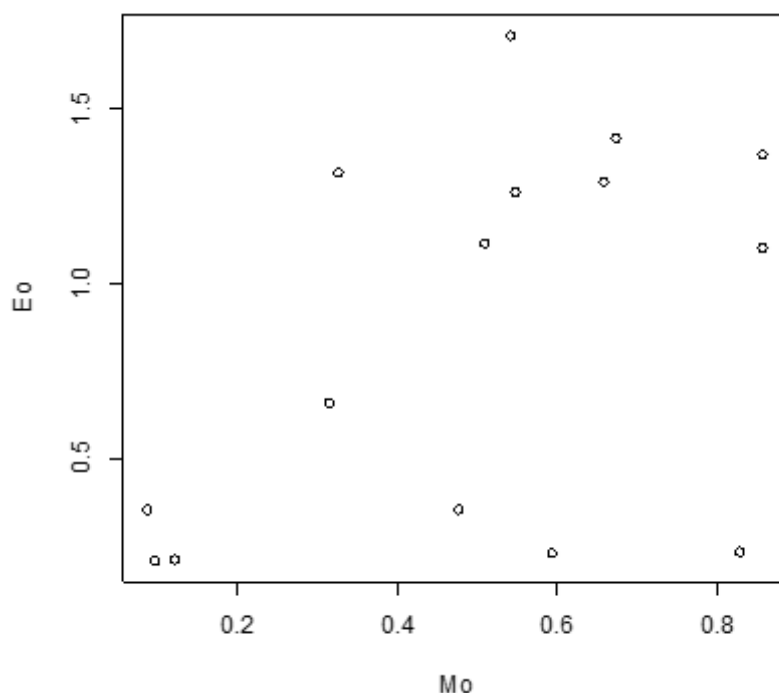
Passamos agora para as correlações da aplicação dos onsets, cujos resultados podem ser visualizados na tabela abaixo:

Tabela 05 - Correlações da aplicação dos onsets

$cor(M_O, E_{O\text{ razão}})$	0.45
$cor(M_O, E_{O\text{ log}})$	0.43

Dessa vez observamos a existência de uma relação mais linear, na direção positiva, entre as previsões do modelo e os dados de nosso experimento. Vejamos também a distribuição dos dados:

Gráfico 05 – Distribuição de $E_{O\text{ razão}}$ em relação a M_O



Utilizando os valores da *pontuação* no lugar dos *valores de entropia máxima* para compor M_O a correlação atingiu os valores de 0.53 (no caso de $E_{O\text{ razão}}$) e 0.51 (no caso de $E_{O\text{ log}}$). Observa-se, portanto, um nível de acurácia consideravelmente maior que aquele obtido para a aplicação completa. Entretanto, esses valores ainda estão bastante aquém dos resultados encontrados por Hayes e Wilson (2008, p.401) em uma aplicação realizada com os onsets do inglês, no qual a correlação chegou a 0.946.

Acreditamos que uma das explicações mais promissoras para tal discrepância esteja na diferença entre os métodos utilizados para obtenção dos dados experimentais. Enquanto utilizamos um experimento de estimação de magnitude aplicado à fonotática (cujos procedimentos foram descritos em 2.2), os autores utilizaram dados de um experimento categórico em aceitabilidade fonotática retirado de Scholes (1966) no qual os participantes respondiam apenas ‘sim’ ou ‘não’ em relação à aceitabilidade das palavras. De posse desses dados, Hayes e Wilson (2008), seguindo procedimentos de Pierrehumbert (1994) e Coleman e Pierrehumbert (1997), tomam a proporção de respostas ‘sim’ em relação ao tamanho completo da amostra como “an indicator of the mean well-formedness intuition of individuals in the population.” (HAYES&WILSON, 2008, p.399). De acordo com os autores, “Frisch, Large, and Pisoni (2000) demonstrate that this method yields scores that are highly correlated with well-formedness ratings on a numerical scale.” Em todo caso, a estimação de magnitude permite maior gradiência

nos julgamentos e não depende de procedimentos indiretos para transformar dados categóricos em estimativas de gradiência na boa-formação. Portanto, é possível que a acurácia do Maxent PL seja de fato menor quando consideramos dados experimentais mais detalhados.

Contudo, precisamos de mais pesquisas uma vez que este é o único trabalho o qual conhecemos que mede a correlação das previsões do Maxent PL com dados experimentais obtidos a partir da utilização da estimação de magnitude. Apesar de Hayes e White (2013) utilizarem a estimação de magnitude em seu experimento fonotático, os autores não estavam preocupados em calcular a acurácia do Maxent PL, e sim em comparar como restrições classificadas como foneticamente naturais e restrições não naturais são julgadas experimentalmente e como o Maxent PL lida com tais categorias, de maneira geral, no aprendizado e nas previsões finais.

Além disso, nossa amostra de 12 participantes é consideravelmente pequena quando olhamos para a população do experimento. Portanto, são necessárias pesquisas em maior escala para que possamos chegar a resultados mais conclusivos. Desejamos, portanto, que o presente trabalho possa servir de base para mais pesquisas com o Maxent PL e outros modelos de Entropia Máxima no Brasil. Esperamos ainda que, quando disponibilizadas ao público, as ferramentas metodológicas construídas para realização das simulações e do experimento, a exemplo dos programas, facilitem novas aplicações do Maxent PL e ajudem em trabalhos computacionais, experimentais e didáticos que delas possam fazer uso.

CONSIDERAÇÕES FINAIS

Neste trabalho, propusemos investigar a correlação Pearson entre as previsões do Maxent PL acerca da fonotática do PB e os dados experimentais. Na introdução, mostramos que a perspectiva gradiente de fonotática adotada pelo modelo se diferencia fundamentalmente de modelos categóricos em linguística formal, o que nos permitiu compreender qual tipo de previsões o modelo produz na forma de uma gramática fonotática. No capítulo I, descrevemos em detalhes os componentes do Maxent PL e como o mesmo aprende uma gramática fonotática a partir de alguns dados de entrada. No capítulo II, aplicamos o modelo ao PB e obtivemos os dados experimentais, obtendo assim as previsões do Maxent PL sobre o PB e os dados empíricos com os quais testaríamos tais previsões.

As limitações da linha de base teórica utilizada na primeira aplicação do modelo nos levaram a realizar uma segunda aplicação, construindo assim dois conjuntos de previsões acerca da fonologia do PB. Ambos os conjuntos foram testados em relação a dados coletados a partir de um experimento de estimação de magnitude no âmbito da aceitabilidade fonotática. No capítulo III, A correlação encontrada para a primeira aplicação foi próxima de 0, revelando a ausência de uma dependência linear entre as previsões e os dados experimentais. A segunda aplicação apresentou uma correlação por volta de 0.5, mostrando a existência de uma dependência linear na direção positiva, mas consideravelmente mais fraca que aquela encontrada para uma aplicação semelhante com o inglês (HAYES&WILSON, p.401).

Ainda no capítulo III, consideramos algumas hipóteses para os resultados obtidos, relacionadas principalmente com alguns parâmetros de aplicação e o método utilizado no experimento. Entretanto, como afirmamos lá, são necessárias mais aplicações do Maxent PL e estudos em maior escala para que possamos chegar a resultados mais conclusivos. Nesta direção, esperamos que, assim que disponibilizadas ao público, as ferramentas metodológicas construídas para realização das simulações e do experimento, a exemplo dos programas, facilitem novas aplicações do Maxent PL e ajudem em trabalhos computacionais, experimentais e didáticos que delas possam fazer uso. Esperamos ainda que o capítulo I tenha deixado claro a generalidade dos modelos estatísticos de Entropia Máxima e as infinitas possibilidades de aplicações a outros domínios da linguística e outras ciências. Assim, que o presente texto possa servir também de introdução a modelos de Entropia Máxima, inspirando a aplicação e criação dos mesmos na ciência brasileira.

REFERÊNCIAS

- BARD, E; ROBERTSON, D; SORACE, A. Magnitude estimation of linguistic acceptability. *Language*, vol. 72, p.32-68, 1996.
- BERGER, A; PIETRA, E; PIETRA, V. A maximum entropy approach to natural language processing. *Computational Linguistics*, vol.22, p.39-71, 1996.
- BISOL, L. (org.). Introdução a estudos da fonologia do português brasileiro, Porto Alegre: EDIPUCRS, 1996.
- CÂMARA JR, J. Estrutura da língua portuguesa. 5. ed. Petrópolis: Vozes, 1975 [1970].
- CHOMSKY, N; HALLE, M. The Sound Pattern of English. Cambridge: MIT Press, 1968.
- CHOMSKY, N. Syntactic Structures. 2 ed. Berlin: Mouton de Gruyter, 2002 [1957].
- _____. Aspects of the Theory of Syntax. Cambridge - Massachusetts: The M.I.T. Press. Massachusetts Institute of Technology, 1965.
- CHISWELL, I. A Course in Formal Languages, Automata and Groups. London: Springer-Verlag, 2009.
- CLARK, S. Maximum Entropy Models for Tagging. Disponível em: https://www.cl.cam.ac.uk/teaching/1213/L101/clark_lectures/lect2.pdf. Cambridge, 2013.
- CLEMENTS, G; HUME, E. The internal organization of speech sounds, In: Goldsmith, J. (org.). *The Handbook of Phonological Theory*, Oxford: Blackwell: 245-306, 1995.
- CLEMENTS, G. Neutral vowels in Hungarian vowel harmony: An autosegmental interpretation. In: KEGL, J; NASH, D; ZAENEN. A. *North Eastern Linguistic Society (NELS)*. 7 ed. Amherst: University of Massachusetts, Graduate Linguistic Student Association, 1976, p.49-64.
- FINATTO, M. Do jornal popular gaúcho ao jornal popular baiano: padrões de frases verbais do Português Popular Escrito - PorPopular Fase 2. Disponível em: http://www.ufrgs.br/textecc/porlexbras/porpopular/files/Relatorio_Site_PorPopular_2016.pdf. Rio Grande do Sul: Universidade Federal do Rio Grande do Sul, 2016.
- FINATTO, M, et al. Lista das 3000 palavras mais frequentes no mini corpus do jornal massa! Disponível em: <http://www.ufrgs.br/textecc/porlexbras/porpopular/massafiles/Veja%20a%20Lista.pdf> Acesso em: 30 de jan, 2017.
- HAHN, L. A Realização da Lateral /l/ no Inglês por Falantes do Português Brasileiro. Pontifícia Universidade Católica do Rio Grande do Sul. Porto Alegre (RS), 2010. (Dissertação de Mestrado).
- HALLE, M. The sound pattern of Russian. The Hague: Mouton, 1959.

HAYES, B; WILSON, C. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry*, vol.30, n.3, p.379-440, 2008.

HAYES, B; WILSON, C; CAPODIECI, F. Manual: Phonotactic learning program. Disponível em: <http://www.linguistics.ucla.edu/people/hayes/Phonotactics/Index.htm#software>. Acesso em: 30 de jan, 2017.
California: UCLA, 2008.

HAYES, B; WILSON, C; CAPODIECI, F. Phonotactic learning program. Disponível em: <http://www.linguistics.ucla.edu/people/hayes/Phonotactics/Index.htm#software>. Acesso em: 30 de jan, 2017.

HAYES, B; WHITE, J. Phonological Naturalness and Phonotactic Learning. *Linguistic Inquiry*, vol.44 n.1, p.45-75, 2013.

HOPCROFT, J; ULLMAN, J; MOTWANI, R. Introduction to automata theory, languages, and computation. Reading: Addison-Wesley, 2001 [1979].

HORA, D. Processo de palatalização das fricativas na Língua Portuguesa. *Revista do GELNE (UFC)*, Fortaleza, n.2, p. 34-36, 1999.

HORA, D. Vocalização da lateral /l/: correlação entre restrições sociais e estruturais. *Scripta (PUCMG)*, vol. 9, 2006, p. 31-46.

JAYNES, E. Probability Theory The Logic of Science. New York: Cambridge University Press, 2003.

KLEIN, D; MANNING, C. Maxent models, conditional estimation, and optimization, without the magic. *Tutorial presented at NAACL-03 and ACL-03*, 2003.

KORNAI, A. Formal Phonology. New York: Garland Publishing, 1995

KRACHT, M. The Mathematics of Language. Berlin: Mouton de Gruyter, 2003.

MCCARTHY, J. A Thematic Guide to Optimality Theory. Cambridge: Cambridge University Press, 2002.

PEREYRON, L. Epêntese vocálica em encontros consonantais mediais por falantes porto – alegrenses de inglês como língua estrangeira. Pontifícia Universidade Católica do Rio Grande do Sul. Porto Alegre (RS), 2008. (Dissertação de Mestrado).

PIETRA, S; PIETRA, V; LAFFERTY, J. Inducing Features of Random Fields. *IEEE TRANSACTIONS PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, v.19, n. 4, p.1-13, 1997

PRESS, W, et al. Numerical recipes in C: The art of scientific computing. Cambridge: Cambridge University Press, 1992.

PRINCE, A; SMOLENSKY, P. Optimality Theory: Constraint interaction in generative grammar. Boulder: Blackwell, 2004 [1993].

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL <https://www.R-project.org/>. 2015

SHANNON, C. A Mathematical Theory of Communication. *The Bell System Technical Journal*, vol.27, p.379-423, 1948.

_____. Saiba mais sobre o jornal massa! on-line. Disponível em: <http://www.ufrgs.br/textecc/porlexbras/porpopular/massafiles/Conheca%20o%20Massa!.pdf>. Acesso em: 30 de jan, 2017.

STANLEY, R. Redundancy rules in phonology. *Language*. vol. 43, p.393–436, 1967.

STERIADE, D. Underspecification and markedness. In: Goldsmith, J. (org.). *The Handbook of Phonological Theory*, Oxford: Blackwell: 114-174, 1995.

STEVENS, S. On the psychophysical law. *Psychological Review*, vol.64 n.3, p.153-181, 1957.

UFFINK, J. *The Constraint Rule of the Maximum Entropy Principle*. Disponível em: <http://dspace.library.uu.nl/bitstream/handle/1874/2648/mep2def.pdf?sequence=1>. Acesso em: 30 de jan. 2017.

ANEXOS

Anexo A1: gramática fonotática da primeira aplicação (completa)

*[-continuant][+word_boundary]	(tier=default)	4,793
*[-voice][+continuant]	(tier=default)	3,002
*[+consonantal,+sonorant][+word_boundary]	(tier=default)	4,518
*[+consonantal,+sonorant][-syllabic]	(tier=default)	4,829
*[-continuant][+consonantal,+labial]	(tier=default)	3,852
*[+consonantal,+labial][-continuant]	(tier=default)	4,175
*[-consonantal,+nasal][-consonantal,+nasal]	(tier=default)	2,268
*[-anterior][+labial]	(tier=default)	3,731
*[-consonantal,+nasal][+low]	(tier=default)	1,5
*[+nasal,-low][+back]	(tier=default)	3,598
*[+consonantal,+labial][+labial]	(tier=default)	3,6
*[+word_boundary][+sonorant,-anterior]	(tier=default)	4,566
*[+low][+low]	(tier=default)	2,831
*[+voice][+continuant,-voice]	(tier=default)	3,417
*[-continuant][-continuant,+voice]	(tier=default)	3,713
*[+low][-round]	(tier=default)	3,2
*[-continuant][+consonantal,+nasal]	(tier=default)	4,132
*[+high,-back][+high,-back]	(tier=default)	4,335
*[+low][+nasal,+back]	(tier=default)	1,882
*[+continuant,+voice][+continuant,+coronal]	(tier=default)	3,191
*[+word_boundary][-consonantal,-syllabic]	(tier=default)	4,861
*[+back][+nasal,+round]	(tier=default)	2,668
*[-high][+nasal,+high]	(tier=default)	3,394
*[+nasal,+high][+approximant]	(tier=default)	3,145
*[^-nasal][-anterior,+lateral]	(tier=default)	4,572
*[+voice][+spread]	(tier=default)	4,034
*[-low,-round][-word_boundary]	(tier=default)	5,584
*[+nasal,+back][-consonantal,-syllabic]	(tier=default)	3,813
*[-continuant,+voice][-approximant,+anterior]	(tier=default)	3,625
*[-back][+low,-back]	(tier=default)	3,527
*[+nasal,-high,-back][+word_boundary]	(tier=default)	4,39

*[-high,-back][-nasal,-high,-back] (tier=default)	3,222
*[+consonantal][-syllabic][-approximant] (tier=default)	4,861
*[-voice,-anterior][+word_boundary] (tier=default)	4,364
*[-consonantal][-high][-high] (tier=default)	3,949
*[-consonantal][+nasal][+word_boundary] (tier=default)	4,197
*[+back][+low,+round] (tier=default)	2,973
*[+consonantal][+sonorant,-syllabic][+word_boundary] (tier=default)	4,489
*[^-nasal,-high,+round][+voice][+word_boundary] (tier=default)	4,456
*[+word_boundary][+continuant,+voice][-syllabic] (tier=default)	4,057
*[-nasal][+nasal,-anterior] (tier=default)	3,908
*[^-nasal,-high,+back][+consonantal,+labial][+word_boundary] (tier=default)	4,113
*[+nasal,-low][-consonantal][^+continuant,-voice,+anterior] (tier=default)	3,83
*[+low][+back,+syllabic][+sonorant] (tier=default)	3,902
*[-consonantal,+nasal][^nasal,-high,+round][+voice] (tier=default)	3,711
*[^-continuant,+coronal][+high,+back][+round,+syllabic] (tier=default)	3,722
*[^-continuant][+nasal,-low][+continuant,+voice] (tier=default)	4,095
*[^-nasal,-high][-consonantal,-syllabic][+round] (tier=default)	3,997
*[-word_boundary][+nasal][+spread] (tier=default)	4,31
*[+continuant,-voice,+coronal][+sonorant,-syllabic][^nasal,-high] (tier=default)	3,963
*[^nasal,-high,-back][+continuant][+low,+round] (tier=default)	3,985
*[^+continuant,-voice,+coronal][+low,+round][+word_boundary] (tier=default)	4,271
*[-high][+syllabic][-sonorant,+labial] (tier=default)	3,981
*[^nasal,-high][+continuant][+low,-back] (tier=default)	3,912
*[^+continuant,-voice,+anterior][+nasal][+consonantal,+approximant] (tier=default)	3,882
*[+high,+syllabic][^sonorant,+anterior][+nasal,+high] (tier=default)	4
*[+syllabic][-high,-back][+word_boundary] (tier=default)	3,973
*[+low,-back][+sonorant,-syllabic][^high,-back] (tier=default)	4,008

Anexo A2: classes naturais na gramática completa

Novel of size 1: +word_boundary #

Novel of size 1: -word_boundary a,an,b,d,e,ee,en,f,g,h,i,in,j,jj,k,l,lh,m,n,nh,o,on,oo,p,q,r,s,t,u,un,v,w,x,z

Novel of size 2: -low,-back e,en,i,in,j

Novel of size 1: +nasal an,en,in,m,n,nh,on,un

Novel of size 3: +nasal,-high,+back an,on

Novel of size 1: +low a,an,ee,oo

Novel of size 3: +high,+back,+syllabic u,un

Novel of size 3: -high,-low,-back e,en

Novel of size 1: -nasal a,e,i,o,u

Novel of size 2: +consonantal,+labial b,f,m,p,v

Novel of size 2: +voice,+coronal d,jj,z

Novel of size 3: +nasal,+high,-back in

Novel of size 2: +nasal,+coronal n,nh

Novel of size 2: -continuant,+voiceb,d,g

Novel of size 2: +nasal,+low an

Novel of size 2: -high,+back a,an,o,on,oo,q

Novel of size 2: -high,+round o,on,oo

Novel of size 2: +continuant,+voice jj,v,z

Novel of size 2: -approximant,+anterior d,n,s,t,z

Novel of size 1: -syllabic b,d,f,g,h,j,jj,k,l,lh,m,n,nh,p,r,s,t,v,w,x,z

Novel of size 2: -nasal,-high a,e,o

Novel of size 2: +approximant,-syllabic j,l,lh,r,w

Novel of size 2: +back,+syllabic a,an,o,on,oo,q,u,un

Novel of size 2: -back,+syllabic e,ee,en,i,in

Novel of size 2: +round,+syllabic o,on,oo,u,un

Novel of size 2: +voice,-anterior jj

Novel of size 2: -low,+back o,on,q,u,un,w

Novel of size 2: -nasal,+back a,o,u

Novel of size 2: +nasal,-low en,in,on,un

Novel of size 1: +sonoranta,an,e,ee,en,i,in,j,l,lh,m,n,nh,o,on,oo,q,r,u,un,w

Novel of size 1: -low e,en,i,in,j,o,on,q,u,un,w

Novel of size 3: -high,-low,+round o,on

Novel of size 2: -continuant,-voice k,p,t
 Novel of size 2: +continuant,-voice f,s,x
 Novel of size 1: +consonantal b,d,f,g,h,jj,k,l,lh,m,n,nh,p,r,s,t,v,x,z
 Novel of size 1: -consonantal a,an,e,ee,en,i,in,j,o,on,oo,q,u,un,w
 Novel of size 2: +consonantal,+approximant l,lh,r
 Novel of size 3: -low,+round,+syllabic o,on,u,un
 Novel of size 1: +voice b,d,g,jj,v,z
 Novel of size 2: -voice,+coronal s,t,x
 Novel of size 1: -anterior jj,lh,nh,r,x
 Novel of size 2: -nasal,+low a
 Novel of size 3: -continuant,+voice,+labial b
 Novel of size 3: -continuant,+voice,+coronal d
 Novel of size 1: -sonorant b,d,f,g,h,jj,k,p,s,t,v,x,z
 Novel of size 3: -nasal,-high,-back e
 Novel of size 3: +continuant,-voice,+labial f
 Novel of size 2: +voice,+dorsal g
 Novel of size 1: +spread h
 Novel of size 2: -low,+syllabic e,en,i,in,o,on,q,u,un
 Novel of size 3: -nasal,+high,-back i
 Novel of size 2: -back,-syllabic j
 Novel of size 2: -voice,+dorsal k
 Novel of size 2: +approximant,+anterior l
 Novel of size 2: +nasal,+labial m
 Novel of size 2: +nasal,+anterior n
 Novel of size 2: +low,+back a,an,oo
 Novel of size 3: -nasal,-high,+round o
 Novel of size 3: -continuant,-voice,+labial p
 Novel of size 2: -low,-round q
 Novel of size 1: -lateral r
 Novel of size 2: -approximant,-anterior jj,nh,x
 Novel of size 2: -consonantal,+nasal an,en,in,on,un
 Novel of size 3: +continuant,-voice,+anterior s
 Novel of size 3: -continuant,-voice,+coronal t
 Novel of size 3: -nasal,+high,+back u

Novel of size 3: +continuant,+voice,+labial v
 Novel of size 1: +syllabic a,an,e,ee,en,i,in,o,on,oo,q,u,un
 Novel of size 2: -consonantal,+labial w
 Novel of size 2: +nasal,+round on,un
 Novel of size 2: -voice,-anterior x
 Novel of size 3: +continuant,+voice,+anterior z
 Novel of size 1: -continuant b,d,g,k,p,t
 Novel of size 2: -anterior,+lateral lh
 Novel of size 1: +high i,in,j,u,un,w
 Novel of size 2: -sonorant,+coronal d,jj,s,t,x,z
 Novel of size 2: +sonorant,+coronal l,lh,n,nh,r
 Novel of size 2: +high,+back u,un,w
 Novel of size 3: +nasal,-high,-low en,on
 Novel of size 2: +low,-round a,an
 Novel of size 3: +high,-back,+syllabic i,in
 Novel of size 1: +labial b,f,m,p,v,w
 Novel of size 2: +consonantal,+nasal m,n,nh
 Novel of size 3: +nasal,+high,+back un
 Novel of size 2: +high,+syllabic i,in,u,un
 Novel of size 3: -low,-back,+syllabic e,en,i,in
 Novel of size 1: +anterior d,l,n,s,t,z
 Novel of size 1: -voice f,k,p,s,t,x
 Novel of size 2: +low,-back ee
 Novel of size 2: -approximant,+coronal d,jj,n,nh,s,t,x,z
 Novel of size 2: +continuant,+coronal jj,s,x,z
 Novel of size 2: +nasal,+back an,on,un
 Novel of size 1: -back e,ee,en,i,in,j
 Novel of size 2: +sonorant,-anterior lh,nh,r
 Novel of size 3: +nasal,-high,-back en
 Novel of size 2: -sonorant,+labial b,f,p,v
 Novel of size 2: -high,-back e,ee,en
 Novel of size 1: +approximant a,an,e,ee,en,i,in,j,l,lh,o,on,oo,q,r,u,un,w
 Novel of size 1: +back a,an,o,on,oo,q,u,un,w
 Novel of size 2: -nasal,-low e,i,o,u

Novel of size 1: +round o,on,oo,u,un,w
 Novel of size 2: +nasal,-high an,en,on
 Novel of size 2: -sonorant,-anteriorjj,x
 Novel of size 2: +nasal,-anterior nh
 Novel of size 3: +continuant,+voice,+coronal jj,z
 Novel of size 2: +nasal,+high in,un
 Novel of size 3: -high,-low,+back o,on,q
 Novel of size 2: -sonorant,+anterior d,s,t,z
 Novel of size 2: -high,-low e,en,o,on,q
 Novel of size 3: -low,+back,+syllabic o,on,q,u,un
 Novel of size 2: +nasal,-back en,in
 Novel of size 2: -nasal,-back e,i
 Novel of size 2: +sonorant,-syllabic j,l,lh,m,n,nh,r,w
 Novel of size 3: -nasal,-high,+back a,o
 Novel of size 1: +dorsal g,k
 Novel of size 2: -continuant,+labial b,p
 Novel of size 2: -low,+round o,on,u,un,w
 Novel of size 3: -nasal,-high,-low e,o
 Novel of size 2: -voice,+labial f,p
 Novel of size 2: -continuant,+coronal d,t
 Novel of size 2: +voice,+labial b,v
 Novel of size 2: +sonorant,+anterior l,n
 Novel of size 1: -high a,an,e,ee,en,o,on,oo,q
 Novel of size 2: +continuant,+labial f,v
 Novel of size 2: -nasal,+high i,u
 Novel of size 2: +voice,+anterior d,z
 Novel of size 3: +nasal,-high,+round on
 Novel of size 2: +low,+round oo
 Novel of size 2: -consonantal,-syllabic j,w
 Novel of size 2: +sonorant,+labial m,w
 Novel of size 2: -nasal,+round o,u
 Novel of size 2: -voice,+anterior s,t
 Novel of size 1: +lateral l,lh
 Novel of size 1: +continuant f,h,jj,s,v,x,z

Novel of size 3: +continuant,-voice,+coronal s,x
 Novel of size 2: +continuant,+anterior s,z
 Novel of size 2: +approximant,-anterior lh,r
 Novel of size 2: +consonantal,+sonorant l,lh,m,n,nh,r
 Novel of size 2: +high,-back i,in,j
 Novel of size 1: -approximant b,d,f,g,h,jj,k,m,n,nh,p,s,t,v,x,z
 Novel of size 1: +coronal d,jj,l,lh,n,nh,r,s,t,x,z
 Novel of size 1: -round a,an,q

Anexo B1: gramática fonotática da segunda aplicação (*onsets*)

*[-word_boundary][-approximant] (tier=default)	2,987
*[+voice,+labial][^lateral] (tier=default)	0,72
*[+coronal][^lateral] (tier=default)	1,604
*[+sonorant][-word_boundary] (tier=default)	2,423
*[-word_boundary][-anterior,+lateral] (tier=default)	0,52
*[+continuant,+coronal][-word_boundary] (tier=default)	1,5
*[^-voice][+anterior] (tier=default)	0
*[+continuant,+voice][-word_boundary] (tier=default)	1,411
*[+spread][-word_boundary] (tier=default)	2,327
*[+nasal,-anterior] (tier=default)	0,51
*[-voice,-anterior] (tier=default)	0,645
*[-anterior,+lateral] (tier=default)	1,519
*[^-continuant][-word_boundary] (tier=default)	2,101
*[-sonorant,-anterior] (tier=default)	0,674
*[-continuant,+labial][^lateral] (tier=default)	0,084
*[^-continuant,-voice,+labial][+anterior] (tier=default)	0,609
*[+voice,+coronal][-word_boundary] (tier=default)	1,86
*[-continuant,+voice,+labial] (tier=default)	1,694
*[+voice,+dorsal] (tier=default)	1,052
*[+continuant,+voice,+anterior] (tier=default)	1,416
*[+continuant,-voice,+labial][+word_boundary] (tier=default)	1,482
*[^-voice,+labial][+lateral] (tier=default)	0,416
*[+dorsal][-word_boundary] (tier=default)	0
*[+continuant,+voice,+labial] (tier=default)	1,092
*[+voice][^lateral] (tier=default)	0,568
*[+word_boundary][+approximant,+anterior] (tier=default)	0
*[-voice,+dorsal][-word_boundary] (tier=default)	0,186
*[+spread] (tier=default)	0,902
*[-continuant,-voice,+labial][+word_boundary] (tier=default)	0
*[+voice,-anterior] (tier=default)	0,029
*[+nasal,+anterior] (tier=default)	0
*[+sonorant,+labial] (tier=default)	0,24

*[^+continuant][^+approximant,+anterior] (tier=default) 0
 *[-voice,+dorsal] (tier=default) 0
 *[^+dorsal][+anterior] (tier=default) 0,918
 *[^+sonorant,-anterior] (tier=default) 0,739
 *[^+continuant,-voice,+coronal] (tier=default) 0,456
 *[-continuant][^~lateral] (tier=default) 1,118
 *[^+coronal][~word_boundary] (tier=default) 1,646
 *[^+dorsal][^+approximant,+anterior] (tier=default) 1,239
 *[-voice,+dorsal][^+approximant,+anterior] (tier=default) 0,746
 *[^+voice,+coronal] (tier=default) 0,153
 *[^+continuant][^~lateral] (tier=default) 0,704
 *[^+nasal] (tier=default) 0,393
 *[^~voice][~word_boundary] (tier=default) 2,174
 *[-continuant,-voice,+labial] (tier=default) 1,07
 *[^~continuant][~anterior] (tier=default) 1,187
 *[^~word_boundary][^~lateral] (tier=default) 2,11
 *[^+continuant,-voice,+anterior] (tier=default) 0
 *[^~approximant,-anterior] (tier=default) 1,184
 *[^+nasal,+coronal] (tier=default) 0,152
 *[^+lateral] (tier=default) 0,023
 *[-voice][^+approximant,+anterior] (tier=default) 2,258
 *[^+dorsal] (tier=default) 0,545
 *[^+labial][^+approximant,+anterior] (tier=default) 1,156
 *[^+approximant,+anterior] (tier=default) 0
 *[-voice][^~lateral] (tier=default) 0,972
 *[^~continuant,+labial][~word_boundary] (tier=default) 0,119
 *[^+labial][^~lateral] (tier=default) 0,651
 *[^+word_boundary][+lateral] (tier=default) 0,9
 *[^~voice,+labial][~word_boundary] (tier=default) 0,702
 *[-continuant,-voice][^~lateral] (tier=default) 0,601

Anexo B2: classes naturais na gramática dos *onsets*

Novel of size 1: +word_boundary	#
Novel of size 1: -word_boundary	b,d,f,g,h,jj,k,l,lh,m,n,nh,p,r,s,t,v,x,z
Novel of size 2: +sonorant,+coronal	l,lh,n,nh,r
Novel of size 1: +nasal	m,n,nh
Novel of size 1: +labial	b,f,m,p,v
Novel of size 2: +voice,+coronal	d,jj,z
Novel of size 1: +anterior	d,l,n,s,t,z
Novel of size 1: -voice	f,k,p,s,t,x
Novel of size 2: -approximant,+coronal	d,jj,n,nh,s,t,x,z
Novel of size 2: +continuant,+coronal	jj,s,x,z
Novel of size 2: +sonorant,-anterior	lh,nh,r
Novel of size 2: +nasal,+coronal	n,nh
Novel of size 2: -continuant,+voice	b,d,g
Novel of size 2: -sonorant,+labial	b,f,p,v
Novel of size 2: +continuant,+voice	jj,v,z
Novel of size 2: -approximant,+anterior	d,n,s,t,z
Novel of size 2: -sonorant,-anterior	jj,x
Novel of size 2: +nasal,-anterior	nh
Novel of size 3: +continuant,+voice,+coronal	jj,z
Novel of size 2: +voice,-anterior	jj
Novel of size 2: -sonorant,+anterior	d,s,t,z
Novel of size 2: -continuant,-voice	k,p,t
Novel of size 2: +continuant,-voice	f,s,x
Novel of size 1: +dorsal	g,k
Novel of size 2: -continuant,+labial	b,p
Novel of size 2: -voice,+labial	f,p
Novel of size 2: -continuant,+coronal	d,t
Novel of size 2: +voice,+labial	b,v
Novel of size 1: +approximant	l,lh,r

Novel of size 2: +sonorant,+anterior	l,n
Novel of size 2: +continuant,+labial	f,v
Novel of size 1: +voice	b,d,g,jj,v,z
Novel of size 2: +voice,+anterior	d,z
Novel of size 2: -voice,+coronal	s,t,x
Novel of size 1: -anterior	jj,lh,nh,r,x
Novel of size 3: -continuant,+voice,+labial	b
Novel of size 3: -continuant,+voice,+coronal	d
Novel of size 1: -sonorant	b,d,f,g,h,jj,k,p,s,t,v,x,z
Novel of size 3: +continuant,-voice,+labial	f
Novel of size 2: +voice,+dorsal	g
Novel of size 2: -voice,+anterior	s,t
Novel of size 1: +spread	h
Novel of size 1: +lateral	l,lh
Novel of size 1: +continuant	f,h,jj,s,v,x,z
Novel of size 2: -voice,+dorsal	k
Novel of size 3: +continuant,-voice,+coronal	s,x
Novel of size 2: +approximant,+anterior	l
Novel of size 2: +sonorant,+labial	m
Novel of size 2: +continuant,+anterior	s,z
Novel of size 2: +approximant,-anterior	lh,r
Novel of size 2: +nasal,+anterior	n
Novel of size 1: +sonorant	l,lh,m,n,nh,r
Novel of size 3: -continuant,-voice,+labial	p
Novel of size 1: -lateral	r
Novel of size 2: -approximant,-anterior	jj,nh,x
Novel of size 3: +continuant,-voice,+anterior	s
Novel of size 3: -continuant,-voice,+coronal	t
Novel of size 3: +continuant,+voice,+labial	v
Novel of size 2: -voice,-anterior	x

Novel of size 1: -approximant b,d,f,g,h,jj,k,m,n,nh,p,s,t,v,x,z

Novel of size 3: +continuant,+voice,+anterior z

Novel of size 1: -continuant b,d,g,k,p,t

Novel of size 2: -anterior,+lateral lh

Novel of size 2: -sonorant,+coronal d,jj,s,t,x,z

Novel of size 1: +coronal d,jj,l,lh,n,nh,r,s,t,x,z

APÊNDICES

Apêndice A – O gradiente

No artigo original (HAYES&WILSON, 2008), os autores apresentam o gradiente como sendo dado por:

$$\frac{\partial}{\partial w_i} \log(P(D)) = O(C_i) - E(C_i)$$

Que na nossa notação seria:

$$\frac{\partial}{\partial w_i} \log(P(D)) = Contagem(C_i) - E(C_i)$$

Como será visto mais adiante na dissertação, será necessário levar o gradiente a 0 (ou a um número arbitrariamente pequeno definido como “muito próximo de 0”). Entretanto, como foi notado pelo presente autor, seria impossível tornar os valores das definições acima comensuráveis em certas condições. Consideremos o seguinte espaço amostral finito consistindo de três elementos:

$$\Omega = \{pfata, bsada, lada\}$$

Agora consideremos a restrição *pf e assumamos que a partir de um dado conjunto de dados, acontece o seguinte (abreviamos *Contagem* para *Cont*):

$$Cont(* pf) = 2$$

Em um caso como esse, podemos facilmente verificar que:

Proposição: Para todo valor de $E(*pf)$, $Contagem(*pf) > E(*pf)$

Demonstração: pela definição de $E(*pf)$, temos que:

$$E(*pf) = p(pfata) * pf(pfata) + p(lad) * pf(lad) + p(rah) * pf(rah)$$

Como $*pf(lad) = 0$ e $*pf(rah) = 0$, a equação se reduz a

$$E(*pf) = p(pfata) * pf(pfata)$$

$$E(*pf) = p(pfata) * 1$$

Como a probabilidade é fixada em 1, sabemos então que:

$$E(*pf) \leq 1 < 2 = Contagem(*pf) \quad \square$$

Este exemplo específico mostra que existe pelo menos um caso em que seria impossível levar o gradiente a 0 pela definição do gradiente proposta pelos autores. Demonstramos o caso geral agora:

Proposição: Se $\sum_{x \in \Omega} C(x) > 0$ e $\sum_{x \in \Omega} C(x) < Cont(C)$, então $Cont(C) > E(C)$

Demonstração: Assumindo que

$$\sum_{x \in \Omega} C(x) < Cont(C)$$

e sabendo que

$$0 < \sum_{x \in \Omega} C(x) < Cont(C)$$

segue, por álgebra, que

$$E(C) < Cont(C) \frac{E(C)}{\sum_{x \in \Omega} C(x)}$$

Como $0 \leq E(C) \leq 1$, e $\sum_{x \in \Omega} C(x) \in \mathbb{N} > 0$, não é difícil verificar que

$$Cont(C) \frac{E(C)}{\sum_{x \in \Omega} C(x)} < Cont(C)$$

e, consequentemente, por transitividade,

$$E(C) < Cont(C) \quad \square$$

O que esta proposição nos diz é que se a definição do gradiente fosse aquela utilizada pelos autores, teríamos de limitar os tipos de restrições àquelas que não satisfazem as condições do teorema para podermos obter $Contagem(C_i) - E(C_i) = 0$. Isso significa que para escolher as restrições de uma *gramática fonotática* teríamos que contar o número de vezes que as restrições violam as formas dos dados para então decidir se podemos utiliza-las em um modelo de Entropia Máxima. De fato, não se quer que modelos de Entropia Máxima limitem as restrições que podemos utilizar, então o presente autor concluiu que algum tipo de simplificação estava sendo feita na maneira como a definição do gradiente foi apresentada em Hayes e Wilson (2008).

Tal suspeita foi confirmada através de correspondência eletrônica com um dos autores do Maxent PL, Colin Wilson. O autor confirmou que, na verdade, o gradiente é dado por

$$\frac{\partial}{\partial w_i} \log(P(D)) = contagem(C_i) - |D|E(C_i)$$

ou, equivalentemente, por

$$\frac{\partial}{\partial w_i} \frac{\log(P(D))}{|D|} = E^*(C_i) - E(C_i)$$

Na segunda opção estaríamos então maximizando $\log(P(D))/|D|$, e não apenas $\log(P(D))$. Como já deve ter sido observado, na dissertação utilizamos a primeira opção, escalonando $E(C_i)$ pelo número de pontos nos dados (i.e. ocorrências de palavras). Segundo Colin, por e-mail, a primeira opção é, de fato, aquela utilizada na implementação do Maxent PL em software (HAYES et al., 2008).

Embora a simplificação realizada pelos autores não diga a verdade sobre o gradiente, ela é de fato não é o único exemplo na literatura (MANNING&KLEIN, 2003), provavelmente porque a necessidade de escalonar $E(C_i)$ por $|D|$ seja considerada óbvia o suficiente pelos pesquisadores para ser omitida da notação. Entretanto, tanto por estarmos trabalhando com um tipo de modelo que não é amplamente difundido no Brasil quanto por uma postura pessoal do autor, decidimos explicitar todos os componentes de nossas proposições por mais dispensáveis que fossem em alguns contextos; e, por isto, decidimos esclarecer a questão neste apêndice.

Apêndice B: definição de sílaba do PB e outros componentes

Faremos agora definições que são motivadas por descrições e análises que vêm sendo feitas desde os primeiros trabalhos fonológicos acerca do PB (CÂMARA JR.1972) e que acreditamos formalizar padrões já bem conhecidos nos estudos fonológicos acerca desta língua. Entretanto, o leitor deve estar ciente de que as definições são de pouco valor para uma teoria linguística e serão feitas com o único propósito de tornar preciso o que queremos dizer ao utilizar os termos definidos no decorrer da dissertação.

Utilizaremos os seguintes conjuntos nas definições:

- I é o inventário fonológico do PB, cujos elementos são os fonemas apresentados nos quadros XX e XX da dissertação;
- V é o subconjunto de I contendo todas as vogais (quadro XX);
- C é o subconjunto de I contendo todas as consoantes (quadro XX);
- $O_1 = C \setminus \{w, j\}$;
- $O_2 = \{l, r\}$;
- $C_o = \{s, h, w\}$

Definição: uma *sílaba do PB* é uma *palavra* cujo *comprimento* s está no intervalo $1 \leq s \leq 4$, e, para cada caso, satisfaz as seguintes condições:

- (i) Se $s = 1$, então $a_1 \in V$;
- (ii) Se $s = 2$, então $a_1 \in V$ e $a_2 \in C_o$, ou $a_1 \in O_1$ e $a_2 \in V$;
- (iii) Se $s = 3$, então:
 $a_1 \in O_1$ e $a_2 \in V$ e $a_3 \in C_o$, ou $a_1 \in O_1$ e $a_2 \in O_2$ e $a_3 \in V$;
- (iv) Se $s = 4$, então $a_1 \in O_1$, $a_2 \in O_2$, e $a_3 \in V$, e $a_4 \in C_o$;

Chamamos as vogais de uma *sílaba do PB* de *núcleo*;

Chamamos de *onset* as consoantes C que estão em uma *sílaba do PB* e que pertencem à O_1 ou O_2 ;

Chamamos de *coda* as consoantes C que estão em uma *sílaba do PB* e que pertencem a C_o ;

Apêndice C1: programa para cálculo do espaço total de restrições

O programa não possui interação com o usuário, então para utilizar os valores desejados para N_1 e N_2 basta alterar os valores atribuídos a estas variáveis no script abaixo:

```
parcela = 0
total = 0
N1 = 97
N2 = 90

for n in range (1,5):
    parcela = (N1**n) + n*(N2*(N1**(n-1)))
    total += parcela
    print total

for n in range (1,5):
    print n
```

Apêndice C2: programa para cálculo do valor de Entropia Máxima de uma forma

Para obter o valor desejado, deve-se primeiramente informar ao programa quantas restrições serão utilizadas (se o usuário já souber que uma restrição não viola a forma uma única vez, então já pode retirar a respectiva restrição desta conta), e, em seguida, para cada restrição, deve-se informar seu peso e o número de vezes que viola a forma.

```
import math

conscount = int(raw_input("Digite o numero de restricoes: "));
maxentv = 0.0;

for i in range(conscount):
    weight = float(raw_input("Digite o peso da restricao " + str(i+1) + ": "));
    violcount = int(raw_input("Digite o numero de vezes que a restricao e violada: "));
    maxentv += weight*violcount;

print "Valor de entropia máxima: ",math.exp(-(maxentv));
```


Apêndice E1: 100 primeiras palavras da amostra de dados (grafemas)

de
o
a
e
que
do
da
em
no
com
na
para
um
os
não
é
uma
foi
por
as
ao
dos
se
mais
ele
segundo
como
anos
mas
pelo
também
disse
está
das
pela
ainda
já
foram
são
ser
sua
ano
dia
tem
polícia
nas
até
eu
seu
isso
pessoas
vai
nos
sobre
entre
rio
às
dois

aos
ou
durante
casa
mil
ter
afirmou
três
estava
estado
caso
após
quando
será
presidente
sem
ontem
cidade
mesmo
só
eles
carnaval
ela
banda
pode
há
prefeitura
governo
além
volta
nesta
fazer
bahia
hoje
festa
apenas
dias
rua
depois
delegado
trabalho

Apêndice E2: 100 primeiras palavras da amostra de dados (transc. automática)

di
u
ə
i
ki
du
də
ěj
nu
kõ
nə
parə
ũ
us
nău
ɛ
umə ##erro
foj
poh
as
au
dus
si
majs
eli
segũdu
komu
ănus
mas
pelu
tăběj
disi
efta
das
pelə
ajda ##erro
ʒa
forău
său
seh
suə
ănu
diə
těj
polisiə
nas
atə
eu
seu
isu
pesoas
vaj
nus
sobri
ějtri
hiu
as
dojs

kõtrə
aus
ou
durāti
kazə
miw
teh
afihmou
tres
estavə
estadu
kazu
apɔs
kwādu
sera
prezidēti
sēj
õtēj
sidadi
mezmu
so
elis
kahnavaw
elə
bādə
podɪ ##erro
a
prefejtura
govehnu
alēj
vowtə
neftə
fazeh
baia
ozi
feftə
apenas
dias
hua
depojs
delegadu
trabaʎu ##100

Apêndice E3: 100 primeiras palavras da amostra de dados (versão final)

d i
 u
 q
 i
 k i
 d u
 d a
 en j
 n u
 k on
 n q
 p a r q
 un
 u s
 n an u
 ee
 u m q
 f o j
 p o h
 a s
 a u
 d u s
 s i
 m a j s
 e l i
 s e g un d u
 k o m u
 an n u s
 m a s
 p e l u
 t an b en j
 d i s i
 e x t a
 d a s
 p e l q
 a j d a
 j j a
 f o r an u
 s an u
 s e h
 s u q
 an n u
 d i q
 t en j
 p o l i s i q
 n a s
 a t ee
 e u
 s e u
 i s u
 p e s o a s
 v a j
 n u s
 s o b r i
 en t r i
 h i u
 a s
 d o j s

k o n t r q
a u s
o u
d u r a n t i
k a z q
m i w
t e h
a f i h m o u
t r e s
e x t a v q
e x t a d u
k a z u
a p o o s
k w a n d u
s e r a
p r e z i d e n t i
s e n j
o n t e n j
s i d a d i
m e z m u
s o o
e l i s
k a h n a v a w
e l q
b a n d q
p o d i
a
p r e f e j t u r a
g o v e h n u
a l e n j
v o w t q
n e x t q
f a z e h
b a i q
o j j i
f e x t q
a p e n a s
d i a s
h u q
d e p o j s
d e l e g a d u
t r a b a l h u